Point-Of-Use Water Treatment Systems In Rural Haiti: Human Health And Water Quality Impact Assessment

By
Arun Varghese

M.P.A. Economics and Public Policy
Princeton University, 2001

B.Engg. Chemical Engineering
Bombay University, 1994

SUBMITTED TO THE DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ENGINEERING IN CIVIL AND ENVIRONMENTAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2002

Signature of Author:______________________________________
Department of Civil and Environmental Engineering
May 14, 2002

Certified by:______________________________________
Harold F. Hemond
Professor, Civil and Environmental Engineering
Thesis Supervisor

Accepted by:______________________________________
Oral Buyukozturk
Chairman, Departmental Committee on Graduate Studies

Point-Of-Use Water Treatment Systems In Rural Haiti:

Human Health And Water Quality Impact Assessment

By

Arun Varghese

Submitted to the Department of Civil and Environmental Engineering on May 14, 2002

in partial fulfillment of the requirements for the degree of

Master of Engineering in Civil and Environmental Engineering

ABSTRACT

Unsafe water is a leading cause of death and disease in economically disadvantaged societies. The development of centralised large-scale water treatment and supply systems has proven to be a slow, expensive strategy to provide safe drinking water in many low-income countries. Governments and non-governmental organisations have therefore increasingly been promoting point-of-use water treatment technologies in communities without reliable municipal water supplies. These technologies aim to be low-cost sustainable solutions that rely on filtration, disinfection and safe storage to improve source water quality. This thesis assesses the health and water quality impact of a point-of-use water treatment program being implemented in rural Haiti by a non-governmental organisation. An observational differential study was used to measure health outcomes in 120 families in the village of Dumay. Bivariate and multivariate statistical methods were used to quantify the impact of the water treatment system in reducing the incidence of diarrhea, after controlling for socio-economic differences in the population. The study established that persons with access to the water purification system experienced a five percentage point lower incidence of diarrhea than an equivalent individual without access to the water treatment system. As part of the water quality impact assessment study, the microbial content of source water and stored water in intervention and non-intervention households was measured using membrane filtration tests. Source water in Dumay, especially that drawn from wells, was found to have a low microbial content. Treated water in intervention households was found to be very pure. Untreated water in non-intervention households was found to be significantly more contaminated than source water, suggesting post-collection contamination is a major problem.

Thesis Supervisor: Harold F. Hemond

Title: Professor, Civil and Environmental Engineering

ACKNOWLEDGEMENTS

This research would not exist but for the cooperation and trust of the people of Dumay, Haiti. Over 120 families in this village welcomed me into their homes, shared detailed personal information in a survey that often lasted half an hour, and allowed me to test their drinking water. I am deeply grateful to the gentle, courteous people of Dumay for their hospitality and kindness. Much contemporary research in development derives from studies conducted amongst poor communities in poor countries. I have come to believe that such research often does more to further the careers of the researchers than to improve the lives of the study subjects. I recognise this may be a valid criticism of my work too. I can only hope the research will be of some use to Gift of Water, Inc and possibly to other NGOs promoting point-of-use water treatment.

I would also like to thank:

Jean Remus Beremus, my exceptionally skilled interpreter, helper and guide in Dumay. I could not have hoped for a more gifted colleague in conducting my survey.

Matt Cyr, for his good cheer and for surveying eight households.

Nathan and Wanda Dieudonne in Santos, for their hospitality, for much useful advice and for being so much fun.

Erla, Magda and all the GWI technicians at Dumay, for their ever-smiling support.

Fredo and Ilyerzer, for driving me and for being good friends.

Caroline, for her inspired suggestion of using an oversize syringe when my hand-pump broke.

Phil and Barbara Warwick, for their hospitality in Florida and for permitting me to assess the GWI program.

Daniele Lantagne, for being a constant source of help.

Patrick Gerland, for being an incredibly reliable source of statistics-support.

Professor Victor Chernozhukov, for useful discussions relating to statistical method.

Professor Harold Hemond, for his support and advice.

Table of Contents

List of Tables

1. INTRODUCTION

Unsafe water is a leading cause of death and disease in economically disadvantaged societies. Over 2 million people die every year of waterborne diseases such as cholera, typhoid fever, amoebic dysentery, and other diarrheal diseases. (Mintz *et al.*., 2001). It is estimated that diarrhea killed more children in the last decade alone than the total number of people killed in armed conflict since the second world war. (Wateraid, 2001).

Despite the horrific human cost of unsafe water, advances in water treatment and supply technologies have been slow to diffuse into low-income countries. Since 1990, the number of people without access to clean drinking water has remained nearly constant at approximately 1.1 billion people (WHO/UNICEF/WSSCC, 2000).

Slow economic growth, political instability, social upheavals and poor institutional development in large areas of the developing world suggest that large-scale municipal water treatment and supply systems are unlikely to be established in many urban communities in the near future. In rural areas, large-scale treatment plants may never be economically viable. Governments and non-governmental organisations have therefore increasingly been promoting point-of-use water treatment technologies in communities without reliable municipal water supplies. These technologies aim to be low-cost sustainable solutions that rely on filtration, disinfection and safe storage to improve source water quality. Another type of intervention has focused on the development of clean water sources. Well development programs in rural areas, for instance, are aimed at reducing dependence on contaminated surface water sources. Both types of interventions are often used in parallel when even superior water sources require further treatment.

A large body of literature has researched the efficacy of point-of-use water treatment systems. A number of studies have randomly assigned point-of-use water treatment systems amongst treatment and control groups and performed follow-up analyses of the differential health impact in the two groups. Studies have also examined the acceptability of various treatment devices and community compliance rates with treatment methodologies. These studies have been of value in improving the design, efficiency and reach of existing programs.

The study documented in this thesis examines the health and water quality impact of a point-of-use water treatment program being implemented in rural Haiti by Gift of Water, Inc. (GWI), a Florida-based non-governmental organisation. The study is based on field research conducted in Dumay, Haiti by the author in January 2002. The purpose of this study is to provide GWI with a statistical basis for program evaluation and to make recommendations for program optimisation. This research also makes an original contribution to the literature on point-of-use water treatment by controlling for and assessing (i) the influence of a wide range of socio-economic factors on health outcomes and (ii) the interaction of these socio-economic factors with access to point-of-use treatment in a cross-sectional observational health impact study.

The study had two objectives. The first was to statistically quantify the health impact of the program using an epidemiology study that measured differential health impacts in intervention and non-intervention sample groups after adequately controlling for relevant sources of variance in the two populations. The second objective was to measure the level of microbial contamination of existing water sources in the area, to contrast this with stored water quality in intervention and non-intervention households and to detect trends that explained variance in stored water quality. The thesis combines both elements of the investigation to identify solutions that may lead to improved program efficiency.

2. LITERATURE REVIEW

This section summarises the results of nine papers dealing with point-of-use water treatment systems. The first two papers build a case for the suitability of point-of-use water treatment systems in developing countries and examine the relative cost efficiency of this approach in meeting water supply challenges. Other papers examine the impact and acceptability of point-of-use water treatment systems when adopted in different countries. Many of these papers deal with trial interventions based on the Centre for Disease Control's Safe Water System, a point-of-use water treatment system based on safe storage containers and chemical disinfection. The purpose of this survey is to provide insight into research methods adopted in previous studies.

2.1 The role of low-cost water treatment technologies in development

Mintz *et al.* (2001) argue that relying on time and resource-intensive centralized solutions like piped, treated water will deny safe water to millions of people for decades. The authors support the extensive promotion of point-of-use chemical and solar disinfection systems as the most effective option in enhancing health and contributing to development and productivity.

The authors review the reasons for the slow propagation of municipal scale water and sewage treatment systems in economically disadvantaged countries. The failure of the United Nations' objective of extending adequate water and sanitation to the whole world in the 1980s is attributed to population growth, lack of sufficient funds, inadequate cost recovery, insufficient trained personnel and the continuation of "traditional policies." In particular, little progress was made in extending services to low-income marginalized urban populations and rural areas. The authors question the sustainability of this approach, arguing that providing safe piped water to dispersed populations in rural areas of developing countries would be prohibitively expensive.

In the absence of central water treatment systems, consumers are responsible for treating water to prevent waterborne disease, the authors argue. Treatment by boiling would inactivate viral, parasitic and bacterial pathogens, but it is economically and environmentally unsustainable and provides no residual protection against recontamination.

The paper maintains that point-of-use water treatment by safe and inexpensive chemical disinfectants is a practical and sustainable alternative to boiling. The authors particularly recommend the use of sodium hypochlorite as disinfectant. "In the past five years, several published field trials of hypochlorite for point-of-use water treatment have established that it is acceptable for and effective at improving water quality in a number of settings, and that its use can reduce diarrheal incidence by up to 85%," the paper states.

The limitations of hypochlorite-based disinfectants include inefficacy against viral and parasitic pathogens, and taste and odour problems when used to treat water with excessive amounts of organic material.

The authors caution that the promotion of safe storage practices is integral to the success of point-of-use water treatment. "The risks of diarrhea due to the contamination of (treated) drinking water during household storage… has been repeatedly observed. ... Furthermore, studies have identified drinking water contaminated during collection, transport and storage as a significant route of transmission during epidemics of cholera and dysentery."

The paper also emphasises the importance of behavioural change in achieving sustained reduction in the incidence of diarrheal diseases. The adoption of a new vessel for water storage would imply new expenses and inconveniences for families. Appropriate promotion of these interventions is therefore integral to the success of point-of-use water treatment systems, the paper argues. "Several innovative approaches have been applied to change behviour in the context of programs to promote point-of-use disinfection and safe water storage. These include social marketing, motivational interviewing and ...community mobilization."

2.2 Cost-efficiency of point-of-use water treatment systems

Reiff *et al.* (1996) estimate the capital investment for treated piped water systems at between US$100 and $150 per person served (all US prices). The cost of boiling sufficient water for drinking, processing and cooking of food, dish washing and hand washing (about 40 litres a day for a family of five) is estimated at about US $150 a year.

This estimate does not account for the full environmental cost of ensuing deforestation. By contrast, a chemical disinfectant -based point-of-use water treatment system is estimated to have an annual cost of between US$1.50 and US$4.00, including amortization of capital equipment.

The paper describes the basic elements of a chemical disinfectant-based point-of-use water treatment system. "The recommended intervention is for the households to obtain and utilize one or preferably two suitable water storage containers in which to disinfect and store the essential quantities of water that need to be free of pathogens, with the containers of a design that will protect the contents against recontamination and enable the production and distribution of the water disinfectant to be managed at the local level."

The paper lists the characteristics of a suitable container as being made of a durable material, resistant to impact and oxidation, easy to clean, lightweight and translucent. High-density polyethylene is often the most appropriate material that is readily available, the paper states. The vessel itself should have a volume between 10 and 30 litres so that it is not too heavy, fitted with handles to facilitate lifting and carrying, and a stable base to help prevent overturning. The inlet should be large enough to facilitate easy filling but small enough to prevent the immersion of objects or hands into the water and fitted with a durable screw-on lid, preferably fastened to the container with a cord or chain. A diameter between 6 and 7.5 cm is considered optimal. The container should also have a device for measuring the correct amount of disinfectant to be dosed incorporated into the container or into the flask that contains the disinfectant. A lid and/or a dropper can be designed to serve this purpose. The paper recommends that the container have a faucet, which is resistant to oxidation and impact, closes easily and can achieve a discharge rate of about one litre of water in 15 seconds and a small air-inlet or capped opening that permits the entrance of air as water is being extracted.

A suitable disinfectant, according to the authors, should be reliable and effective in the inactivation of a range of pathogens under the conditions likely to be encountered. It should also provide an adequate residual concentration in the water to assure safe microbial quality throughout the storage period. The disinfectant should not produce deleterious disinfection by-products that make the water unsuitable for human consumption or be aesthetically unacceptable to the consumer. The disinfectant should be reasonably safe for household storage and have an adequate shelf life without loss of potency. It should have an accurate, simple and rapid test for measurement of the disinfectant residual. Finally, the paper states, it should be affordable to the common household. The authors believe that sodium hypochlorite is the disinfectant that best meets these criteria.

The paper estimates the cost of an appropriate container as between US $4 and $6. The cost of transportation can equal the cost of the container, depending on location, highlighting the importance of local manufacture to project sustainability. Typical costs of sodium hypochlorite generators, which work by electrolysing a brine solution, range between $1,500 and $2,000 for units that produce one kilogram of available chlorine in a 24-hour period. "The cost of producing the sodium hypochlorite solution at the community level varies widely from one location to another primarily because of large differences in the cost of salt, electricity, and labour, and to a lesser extent because of differing amortization rates and equipment efficiency," the authors state. However, on average the estimated cost per kilogram of available chlorine produced ranges from about $2.50 to $5, with an average of slightly less than $3. At this average price, a year's supply of 0.5% sodium hypochlorite for a typical family dosing at a rate of 2mg/litre and using 40 litres per day of water for essential purposes is estimated at less than $0.1. The cost of bottling and distributing the hypochlorite in the community raises this to about $1 /year if the flask containing the sodium hypochlorite is reused.

2.3 Field tests of point-of-use water treatment systems in Bolivia

Quick *et al.* (1999) field tested a water quality intervention consisting of point-of-use water disinfection, safe storage and community education in Bolivia in 1994. A total of 127 households in two peri-urban communities were randomized by a public lottery into two groups: one to receive the intervention and the other to serve as a control group. Community health volunteers distributed one container of disinfectant and two special vessels to each intervention household and explained how to treat and store water with these products. "Once a week, community health volunteers distributed containers with freshly prepared disinfectant to each intervention household, removed old containers and used the labels on the special vessels to reinforce messages about proper use of disinfectant and vessels and remind participants of different applications for treated water," the paper states. The primary drinking water source for these households was shallow uncovered household wells. Beginning two months after distribution and extending for a period of five months, a specially-trained health worker made weekly visits to all households to obtain information about all household cases of diarrhea, defined as 3 or more loose or watery stools in 24 hours, with onset in the preceding 7 days. The study reports that there were no statistically significant differences between excluded and participant households in demographic characteristics, sanitary conditions, water handling practices or baseline *E. coli* colony counts in either well or storage water, the study reports.

The study found that intervention households had fewer reported episodes of diarrhea than did members of control households, in all age groups. The protective effect was strongest for infants (age0-1), among whom the reduction in incidence was 53% (p=0.02) and for children 5-14 years old, among whom the reduction was 59% (p=0.01). Reductions in the mean number of diarrhea episodes for persons in the age groups 1-4 years and >15 years did not reach statistical significance. Univariate generalized estimating equations analysis of potential risk factors for diarrhea among individuals revealed that diarrhea risk was less for older persons and for individuals who belonged to the intervention households. Diarrhea risk was greater for males. Diarrhea risk tended to be less for individuals living in households with a latrine in active use, but this result was not statistically significant. The authors then constructed a model that included the statistically significant and borderline significant risk factors from the univariate analysis. Multivariate generalized estimating equations analysis of the model showed that belonging to an intervention household and older age was independently associated with having fewer episodes of diarrhea. Male sex was independently associated with having a higher incidence of diarrhea. Interactions were tested between the independently associated variables and none was found to be significant.

2. 4 Diarrheal disease transmission through water distribution systems

Semenza *et al.* (1998) conducted a randomized intervention trial in Nukus, Uzbekistan to generate epidemiological data to assist policy formulation in the prioritization of competing water treatment technologies. The authors interviewed residents of 240 households, 120 with and 120 without access to municipal piped water. Residents of 62 households without piped water were trained to chlorinate their drinking water at home in a narrow-necked water container with a spout. All study subjects (1,583 individuals) were monitored biweekly for self-reported diarrheal illness over a period of 9.5 weeks. The home chlorination intervention group had the lowest diarrheal rate (28.8/1,000 subjects/month) despite lack of access to piped water in their homes. Compared with the two groups that did not receive the intervention, this rate was one-sixth that of the group with no piped water (179/1,000 subjects/month) and one-third that of the households with piped water (75.5/1,000 subjects/month). More than 30% of the households with piped water lacked detectable levels of chlorine residuals in their drinking water despite two- stage chlorination of the source water and were at increased risk of diarrhea. Forty-two percent of these municipal users reported that water pressure had been intermittent within the previous two days.

The authors conclude that the dramatic reduction in diarrheal rates in the home-chlorination intervention group indicates that a large proportion of diarrheal diseases in Nukus are water-borne. The home chlorination group had less diarrhea than the group with piped water, implicating the distribution system as a source of disease transmission. The authors suggest that their results support the hypothesis that diarrhea in the piped water group could be attributed to cross-contamination between the municipal water supply and sewers, due to leaky pipes and lack of water pressure.

2.5 Field tests on point-of-use water treatment systems in Zambia

Quick *et al.* (1998) conducted field tests of a water quality intervention that consisted of point-of-use chemical disinfectant-based water treatment, safe storage and community education in Kitwe, Zambia in 1998. A total of 166 intervention households were randomly selected from one community and 94 control households from another. Baseline surveys were first conducted and the intervention was thereafter distributed. Weekly active diarrhea surveillance, biweekly water testing and a follow-up survey were conducted over a period of 3 months. There were no statistically significant demographic or socio-economic differences between the treatment and control groups.

During the 5-week baseline period, 103 episodes of diarrhea were detected; 64 episodes (6.4%) among 1,003 persons in intervention households, and 39 episodes (6.7%) among 578 episodes in control households. This difference in incidence rates was not statistically significant.

Over the 8-week period following the launch of the intervention, 22 episodes of diarrhea (2.2%) were reported from 1,003 persons in the treatment group, compared to 28 episodes (4.8%) in the control group. Univariate generalized estimating equations revealed a statistically significant difference in household diarrhea rates (estimated OR 0.53) and individual diarrhea rates (estimated OR 0.52) between the intervention and control groups in the post-launch period. Multivariate generalized estimating equation analysis did not reveal an independent association between diarrhea incidence and variables other than the statistically significant difference observed between intervention and control households. The authors note that a major limitation in their study was an inability to randomize their population sample into treatment and control groups within the same community.

2.6 Impact of point-of-use water treatment on cholera prevention

Reller *et al.* (2001) investigated the risk factors for cholera transmission through a case-control study in Fort-Dauphin, Madagascar in 2001. The Cooperative for Assistance and Relief Everywhere (CARE) had been implementing a household-based safe water intervention in that city since 1999. Cases were selected from 113 patients registered at a cholera treatment ward of a city hospital. For each case, 2 age-, sex- and neighbourhood-matched control subjects were selected from households free of diarrhea during the outbreak. Patients and control subjects were interviewed about beverages and foods consumed in the five days before the patient's illness.

The authors performed univariate and multivariate analysis including conditional logistic regression to determine independent risk factors for infection. The study found that patients were more likely than control subjects to have drunk untreated water from any source (matched odds ratio=5; 95% confidence interval = (1.3, 25.4). Drinking water from a household tap was protective against cholera (OR=0.1; 95% confidence interval (0.0,0.6). Using the chlorine-based disinfectant and boiling water was found to be similarly protective. Illness was not found to be associated with consuming lemonade, unwashed produce, or foods and beverages from street vendors. Using soap to wash hands was protective against illness. In a multivariate model that controlled for the differences in diet between patients and control subjects, illness was independently associated with consuming untreated water. Although the protective effect of the disinfectant being marketed by CARE was not found to be statistically significant because of small numbers, the estimated effect was highly significant (OR=0.1). The protective effect of consuming chicken, eggs or milk -- all expensive products in Fort Dauphin -- was likely a surrogate for relatively higher socioeconomic status.

The study concludes that untreated water was the principal vehicle of epidemic cholera in Fort Dauphin. The community was found to be at risk for waterborne illness despite having access to piped water. Possible reasons for increased risk included inconsistent chlorination of municipal water and domestic storage in wide-mouthed buckets, which permitted hands to touch and contaminate stored drinking water.

2.7 Narrow-mouthed water storage vessels in a Bolivian community

Venczel *et al.* (1996) evaluated the acceptance and impact of an intervention consisting of a narrow-mouthed, plastic water storage vessel and 5% calcium hypochlorite solution for home disinfection of stored water in a Bolivian Aymara Indian community at risk for cholera. The authors systematically selected 42 households from a group of 55 community volunteers in El Alto. These households were randomized into three groups: one group of families received the special vessels and chlorine (group A); another group just the special vessels (group B); the third group served as a control (group C). A baseline study revealed that each of the 42 families in the study obtained water from a household well; faecal coliform bacteria were found in water from 39 of 42 (93%) wells and 33 of 42 (79%) usual water storage vessels. After the intervention, water samples collected every three weeks from group A intervention vessels had lower geometric mean faecal coliform colony counts and lower geometric mean *E. coli* colony counts than water from group B or C vessels. The study also found that the special vessels and chlorine were well accepted in the community and continued to be used for at least six months.

2.8 Acceptability of point-of-use water treatment interventions in rural Kenya

Makutsa *et al.* (2001) studied the acceptability of chemical-disinfectant based point-of-use water treatment and safe storage systems in a rural area of western Kenya. A baseline survey revealed that most participants rarely treated water despite believing that contaminated drinking water was the main cause of diarrhea in their villages. Survey and focus group data revealed that 91% of households stored drinking water in open-mouthed clay pots and that they preferred these pots to plastic jerry cans. A non-governmental organization thereafter proceeded to implement the Safe Water System (a point-of-use treatment plan promoted by the Centres for Disease Control) with partial cost recovery and with clay pots modified to local preferences. The paper proceeds to describe the various social marketing techniques used to effect behavioural change in the communities. Six months after the introduction of the intervention, the authors monitored product adoption in a random sample of 20% of households in 12 project villages. Water stored in 58 of 173 households (33%) had detectable levels of free chlorine residuals, indicating use of the disinfectant, and 18.5% of the households were using the modified clay pots for water storage. The authors find that the adoption rate for chlorination was substantially higher than the rates measured in urban projects in other countries. This finding was found remarkable because of the impoverishment of the project communities and was attributed to the perceived need for water treatment to prevent diarrhea, interest in chemical disinfection and willingness to pay for it. The adoption rate of the modified vessel was also considered satisfactory, in view of the low purchasing power of the community.

2.9 Efficacy of safe storage systems in refugee populations in Malawi

Les Roberts *et al.* (2001) assessed the ability of a water container with a cover and spout to prevent household contamination of water in a Malawian refugee camp. A randomized trial was conducted in a refugee population that had experienced repeated outbreaks of cholera and diarrhea and where contamination of water in the home was found to be a significant cause of cholera. Four hundred Mozambican refugee households were systematically identified and followed over a four-month period. One fourth of the households were randomly assigned to exclusively use the improved container for water collection.

Wells in the vicinity of the sample population were visited beginning a week after the final distribution of the improved buckets. As numbered buckets were filled at the wells, the bucket number, the time of filling, the type of bucket, and the sex and approximate age of the water collector were recorded. A lag time was systematically assigned to each bucket after which the investigators would visit the household and sample water from that particular bucket. This water was tested for faecal coliform and chlorine residuals. To assess the source of the initial bacterial contamination in buckets, the level of contamination on the interior surfaces of buckets and on the hands of women was measured. All study households were visited twice per week and the inhabitants were asked if anyone had experienced diarrhea and if the households possessed soap. The surveillance lasted from late January through the end of May 1993.

The study found that faecal coliform values were 53% lower in the improved buckets than in the ration buckets. When the investigators added 2.5 mg/l chlorine to the buckets, microbial contamination was virtually eliminated for the first four hours but was considerable after 6 hours. The fingers of 10 women arriving at a well were rinsed in 125 ml of well water for 5 seconds. The average rinse recovered more than 2000 faecal coliforms. The buckets of the same women were rinsed with 125 ml of clean well water. On average, the buckets yielded over 300 faecal coliforms.

The study also found that the 310 study participants whose homes received improved buckets experienced 60 episodes of diarrhea in the five months of the study, an attack rate of 44.5 episodes/1000/month. The 850 individuals in control households who remained throughout the study experienced 207 diarrheal episodes for an attack rate of 48.6 episodes/1000/month. This difference was not statistically significant.

For under five-year-olds, there was a 31% reduction in diarrheal incidence associated with use of the bucket ($p=0.06$). Poisson regression models found that among all age groups families which possessed a greater number of huts, an increased number of buckets, and the presence of a latrine were all associated with less diarrhea. Among these, only the association with the improved bucket was statistically significant.

The authors make a particular note that bucket washing practices can contribute to water contamination. "Women often queue for hours in order to fill their buckets in Nyamithuthu camp. Almost always, as a woman steps up to the pump for her turn, she will rinse her bucket with a small amount of water and rub her hand around the inside of the pail. This attempt to be hygienic is almost certainly responsible for the dramatic contamination of water in the standard control buckets between the time when it flowed coliform free from the pump outlet and seconds later when the samples were taken. Educational messages should reinforce that generally human hands are much more contaminated than dry surfaces."

3.BACKGROUND

3.1 Haiti

3.1.1 Geography: The Republic of Haiti is a small, poor, densely-populated country that occupies the rocky western third of the Island of Hispaniola, located between the Caribbean Sea and the North Atlantic Ocean. Its neighbour to the east, the Dominican Republic, occupies the rest of the island.

3.1.2 Human Development Indicators: Haiti ranked a low 134 on the UNDP's human development index, a composite indicator of human progress, in 2001 (UNDP, 2001). Life expectancy at birth in Haiti is 52 years, much lower than the regional average of 70. Adult literacy is only 49%. The infant mortality rate per 1000 live births is 70, more than twice the regional average. The under-5 mortality rate per 1000 live births is 129. Malnutrition affects 28% of children under 5 years. Only 46% of the population has access to an improved water source and only 28% use adequate sanitation facilities. Haiti is one of the most densely populated countries in the region, with a fertility rate of 4.8 compared to the regional average of 2.8. Haiti’s population was estimated at 8 million in 2000.

3.1.3 Economy: Haiti is the one of the poorest countries in the Western Hemisphere, with a gross national income per capita of US $ 480 (Atlas method) in 2000 (World Bank, 2001). This compares unfavourably with even sub-Saharan countries and is much lower than Haiti’s neighbours in Latin America and the Caribbean. The total gross domestic product was only $4.3 billion in 1999. The economy steadily stagnated in the past decade with a per capita GDP growth of -3.4% in 1990-99. The agricultural sector accounted for a little under 30% of gross national product, with industry and services accounting for 21% and 49% respectively in 2000.

3.1.4 Climate and Environment: Haiti experiences tropical climatic conditions except in the semi-arid east, where mountains cut off the trade winds. The terrain is mostly rough and mountainous. The island lies in the middle of the hurricane belt and is vulnerable to severe storms from June to October. The country has suffered extensive deforestation and soil erosion, particularly after US sanctions in the 1990s. Much of the remaining forest is being cleared for agriculture and fuel.

3.1.5 History and Politics: Christopher Columbus, a Spanish explorer of Italian descent, first landed on Hispaniola in 1492, then a lush, densely-populated island inhabited by an estimated three million Arawak Indians who referred to their home as Hayti, or mountainous land. Attracted by the fertility of the island, Spain, France and Britain competed to control the natural resources of Hispaniola to fuel their growing industrial economies. By the middle of the 17th century, Haiti had become a French colony while the Spanish controlled the eastern part of the island. Barbaric colonialist practices devastated the indigenous population, which was soon driven to near extinction. As was their practice, the European colonists looked to Africa as a free source of labour. Thousands of slaves were abducted from West Africa to work in horrifying conditions in the sugarcane, cotton and coffee plantations that were at the heart of European prosperity. In 1792, the brutalized slave population initiated a heroic revolt against the French colonists and by 1804 the island became the first independent black republic in the world.

Despite its sovereign status, Haiti continued to be subject to unsolicited foreign intervention. In 1915, the United States of America invaded the country to "stabilize" it. After US withdrawal in 1934, power was consolidated in the hands of an elite mulatto class sympathetic to US interests. The bulk of the country remained impoverished, deprived of genuine political empowerment and vulnerable to human rights abuses. In 1990, Jean-Bertrand Aristide, a charismatic Roman Catholic priest, swept a presidential election but was soon overthrown in a military coup. In September 1994, a US-led intervention reinstated Aristide. Since regaining power, Aristide has been reluctant to enact the political and economic reform that he was expected to, leading to a souring of relations between his government and that of the United States and a de facto trade embargo against Haiti. The Aristide regime has acquired a harsh dictatorial character and is increasingly resented both within Haiti and internationally for its political repression and its inability to relieve the country's economic stagnation.

3.1.6 Religion: Most of the country professes the Roman Catholic faith. The Protestant church has been attracting more followers in recent years as a result of missionary activities. Much of the Christian population continues to observe voodoo traditions of African origin. When describing the country's religious make-up, observers are wont to quip that "the country is 80% Catholic, 20% Protestant and 100% Voodoo."

3.1.7 Intestinal Infectious Diseases: Diarrheal disease was the leading cause of illness and death in children under 5 years of age in the 1990s (PAHO, 1998). The incidence of diarrhea in the general population was as high as 47.7%, according to health surveys conducted between 1987 and 1994. Typhoid is endemic in Haiti. It ranked as the fifth leading cause of hospitalization during some periods of the 1990s.

3.2 Gift of Water, Inc.: Gift of Water, Inc.(GWI) is a public charity based in Brevard County, Florida. Its mission is "to provide clean drinking water and community development to the impoverished people of developing countries through the use of home-based, appropriate technology water purifiers." (www.giftofwater.org)

GWI initiated its activities in Haiti in 1995 and currently operates water treatment programs in seven different communities across Haiti. The charity works with various church-based organizations and aims to "meet not only physical but also spiritual needs of the disadvantaged" although it "strives to be undiscriminating in the communities it helps." GWI has distributed approximately 3,000 filtration systems amongst these seven communities.

3.3 Dumay

The site of GWI's first water treatment intervention in Haiti, Dumay is a cluster of villages located approximately 15 kilometres south of Port-au-Prince. The study documented in this thesis was conducted mainly in Dumay. About 5% of the data was gathered from two small villages called Bonnette and Beauge, which are located approximately five kilometres outside of Dumay.

Dumay is a flat plain ringed by hills that are largely bare of vegetation. Deforestation and erosion have exposed stony white swathes in some elevated areas. The plain below is still green and mostly cultivated. Several springs and a small river flow through the land. Some of these surface water sources are channeled for irrigation. Principal crops grown in the area include sugarcane, sweet potatoes, corn, beans, tomatoes and vegetables.

The villages are spread over an area of approximately 15 square kilometres, mostly comprising cultivated fields. The roads connecting the village clusters are unpaved. Private operators who drive jeeps called tap-taps provide intermittent public transport.

The villages are named Haut Campeche, Bas Campeche, Celicourt, Temoulin, Tijardin, Barriere Rouge, Lorial, Barriere L'Hopital, Denis, Maroseau, Jean Mary, La Hatte, Liziere, Delmas, Turbe, Jonc, Drouillard, Barron, Gamant, Boiscabrit, Pierroux Douceur, Pont Dumay, Galette Dumay, Pernier, Carrefour Pernier, Timoulin, Bambour, Galette Drouillard, Terresalee, Digneron, Rocheblanche, Coupont, Guedon, Duval Amboise, Laferme, Michaud, Haut Cottard, Laferronnee, Trois Rigeoles and Noailles. These names correspond to different areas of what may be considered one large village.

There is no public hospital in Dumay, but some churches hold occasional medical clinics. There are numerous churches in the village. Four major schools affiliated to local parishes serve the area. Much economic activity in Dumay is agriculture-related. Nearly all the households surveyed worked the land, either in a share-cropping arrangement or as agricultural labourers. A few people owned land and hired labour to cultivate it. Goods are traded at a weekly market. A number of women worked part-time as vendors. A variety of farm animals are raised including cows, goats, sheep, pigs, chickens, ducks, turkeys, donkeys and even horses.

Although agriculture accounted for most employment in Dumay, there was considerable divergence in household wealth and quality of housing. This may be attributed to skewed land distribution and expatriate incomes from relatives working in the United States in some families.

The community did not show large divergence in educational attainment. The average family education deficit, defined as the difference between ideal and actual educational attainment for age, was 6.5 years with a standard deviation of 2.8 years. The study found only 4 university graduates amongst 841 individuals.

The survey indicated that the majority of the population in Dumay was Protestant. About 60% of the survey population was Protestant and a little over 30% was Catholic. More detail on the survey population is available in section 4.8.1.

3.4 The GWI Filter

GWI's current chlorine-based purifier design comprises two detachable 19-litre plastic buckets connected by a check-valve. (Mohamed, 2000). Users fill the top bucket with water, add a 5 ml dose of 5.25% sodium hypochlorite solution, and allow the water to stand for 30 minutes. This contact time with chlorine is expected to kill bacteria and viruses in the water. At the end of 30 minutes, the top bucket is lifted onto the check valve fitted to the bottom bucket, which starts flow into the lower bucket. Water flows through a polypropylene sediment filter in the top bucket and into the bottom bucket through a granular activated carbon (GAC) filter. The GAC removes the chlorine and many other chemicals that might be present in the water. A spigot on the bottom allows users to draw clean water directly from the purifier. Five drops of residual chlorine added to the bottom bucket prevent pathogen regrowth during storage.

The filters are currently manufactured in the United States and assembled in Haiti. The filters cost US $15 and cost recovery is currently US $2.

4. HEALTH IMPACT STUDY

4.1 Objectives

Previous studies have established the efficacy of the GWI filter in purifying water under laboratory conditions. Studies have also investigated participant compliance rates with recommended filter use practices (www.giftofwater.org). With an effective filter and participant compliance, one would infer that populations using the filter should experience a lower incidence of water-borne disease than populations not covered by the intervention. The first portion of this thesis attempts to assess how much healthier filter-using families are compared to those without filters and to what extent this improved health should be attributed to filter use. The study specifically sought to analyze the following issues:

- Are the treatment and control groups identical in all respects except for their use of filters? Stated differently, is there selection bias in the distribution of the filters?
- Is there a statistically significant difference in the incidence of water-borne disease between the treatment group (families using filters) and control group (families without filters)?
- What is the variance in health outcomes within each group? How much of this variance is explained by socio-economic, environmental and behavioural factors? How much is due to filter use?

A systematic investigation of these questions is expected to help quantify program impact and facilitate program optimization.

4.2 Survey Design

Treatment and control groups in an epidemiology study should ideally be either (i) homogeneous in all aspects except in coverage by the specific intervention being investigated or (ii) identically heterogeneous in all aspects. If these conditions are met, a causative link may more credibly be drawn from the differential outcomes observed between the two groups. For instance, if the treatment group showed two fewer incidences of diarrhea per person per year than the control group at a statistically significant level, one could hypothesize a credible causative link between filter use and reduced incidence of diarrhea.

In some of the studies described in the literature review, researchers randomized the intervention of interest to avoid having to consider problems caused by differences in population characteristics. My survey was different from these studies in that it was an observational study and not a randomized one. In other words, I chose to study the program as it existed in Dumay rather than create a laboratory-like situation. It would not have been possible to conduct a randomized trial, given time and resource constraints. Moreover, I believe an observational study was instructive in understanding the impact of the program as it operated, rather than how it might have operated with perfect controls on population characteristics. I therefore needed to gather data on an additional vector of socio-economic and behavioural data that were likely to have an impact on the health outcome being studied. A failure to control for variance in these related factors would possibly bias the results, either giving too little or too much weight to the filter in explaining differential health outcomes between the groups. (Omitted variable bias occurs when the omitted variable is correlated with the intervention whose impact is being investigated). Apart from its non-interventional nature, an observational study had the advantage of quantifying the impact and interaction of other variables on the outcome being investigated. Such information can in itself be vital to understanding the impact of the treatment. However, a major disadvantage of an observational study is the need for large amounts of data before statistically significant trends can be extracted from the analysis. The higher the number of covariates, the greater the sample size needed. Problems of inadequate sample size have restricted the scope of my analysis as will later be apparent. However, these problems were fairly minor.

It is important to emphasize that unlike some of the time-series studies in the literature, this was a cross-sectional survey that polled a randomly selected set of respondents just once. The inability to conduct follow-up surveys is attributable to time and resource constraints. However, I believe a cross-sectional approach is valuable for the snapshot it provides of the differential health outcomes experienced by intervention and non-intervention households in Dumay between December 2001 and January 2002.

4.3 Sampling Methodology

A total of 120 families were surveyed in Dumay, 62 of whom owned a GWI filter. The remaining 58 families were not covered by the program. I shall refer to the sample that used the filter as intervention households and to the remaining sample population as non-intervention households. Intervention households were randomly selected from GWI's program records. The GWI program distributes its filters by a means of a promotion drive in each circuit at the end of which interested families are required to visit the GWI program office and apply for a filter. Each circuit, corresponding to a village unit, was sampled in proportion to its representation in GWI's program. I was unable to perfectly randomize non-intervention households because I lacked access to census or land records. I therefore tried to pick houses at random, but followed no particular randomization algorithm. I tried to maintain a one-to-one ratio of filter to non-filter houses in each circuit, but this was not possible in some of the more distant circuits, which I was able to visit only once. I believe I obtained a fairly representative sample of the non-filter households, but I suspect that I may have undersampled the relatively more wealthy non-filter houses. I consider the inability to randomize non-filter households one of the most serious drawbacks of this survey. Chosen families were surveyed regardless of whether or not they had stored water in their homes.

4.4 Survey Implementation

One respondent in each family was administered a survey that solicited information on the health of each family member in the past month, particularly diarrheal incidence. The respondent was in most cases the mother or grandmother of the family. In addition to health information, the respondent was polled for socio-economic data, described in 3.3.4 below.

The survey was conducted in Creole, the language most widely spoken in rural Haiti, by Jean Remus Beremius, a senior GWI technician, and myself. (Eight households were polled by Matt Cyr, a GWI employee fluent in Creole.) I noted down the survey responses after Remus translated them into English.

The accuracy of some responses may have been compromised by the reluctance of respondents to speak openly of their illnesses in front of Remus, a well-regarded community member. The incidence of HIV seropositivity is around 5% in rural Haiti and families may have feared that admitting to illness might be construed as evidence of HIV. Each family was requested to be as accurate in their responses as possible. My own subjective sense is that the responses were well considered and truthful. Many more people reported incidence of fever than diarrhea. Given that fever is well known to be a symptom of HIV/AIDS in the community, it seems unlikely that a family would admit to fever and not diarrhea to prevent an adverse assumption of HIV infection.

Some families may also have construed the survey as a test to determine whether they ought to receive a filter. As a consequence, respondents may have tried to provide "ideal" responses to questions on behavioural issues and possibly even on health status.

Another data quality problem relates to the accuracy of health data gathered from one respondent about other family members, particularly for a duration of one month. It is quite possible respondents did not have perfect knowledge of diarrheal episodes experienced by all members of the family, or that they did not precisely remember whether the episode had occurred in the past month. Given the proximate living conditions in Dumay, where families of 8 or more persons often shared just one or two rooms, I believe that information gathered from mothers and grandmothers about the health status of their children would be accurate. I also cross-checked each reported incidence of sickness with the respondent to confirm that it had occurred in the past month**.**

4.5 Survey Data

I attempted to gather information on all measurable variables that were likely correlated with health outcomes. In particular, I tried to include measures of variables that are likely correlated with both health outcomes and inclusion in the GWI program. For instances, I anticipated that income is a variable likely to be correlated with inclusion in the program (wealthier people are more likely to have filters) as well as health outcomes (wealthier people are likely to be healthier). Excluding a control for income would thus exaggerate the impact of having the filter on health outcomes.

The survey gathered information on the following variables and groups of variables:

1. Health Outcomes: (i) A binary variable for whether or not each member of the family had experienced diarrhea (defined as three or more loose stools in a 24-hour period) in the past month. (ii) A binary variable for whether or not each member of the family had experienced fever (as defined by the family) in the past month.
2. Filter: a binary variable for whether or not the family used a filter.
3. Geographic Location: A category variable for the circuit number assigned by GWI to each area, which roughly corresponds to a village unit.
4. Household Size: A continuous variable for the number of persons living in the surveyed house.
4. Source of Water: A category variable for source of water. Most families used one of two main sources. (i) Piped spring water capped at the source of springs in the surrounding hills and available at common village taps. (ii) Hand-pumped tube wells constructed by non-governmental organisations.
5. Sanitation Facilities: A category variable that measured whether a family used (i) a private bathroom (ii) a common bathroom or (iii) no sanitation facilities.
6. Quality of Housing: A category variable that classified houses as : (i) Earthern walls and floor, corrugated iron roof (ii) Earthen walls, cement floor, corrugated iron roof (iii) Cement walls, floor and roof, unpainted, unfinished fittings (iv) Cement walls, floor and roof; partially finished fittings (v) Completed concrete structure with modern fittings.
7. Rooms: A continuous variable for the number of rooms in the house
8. Electricity: A category variable for (i) No electricity (ii) Illegal Connection (iii) Legal Connection (iv) Generator

9. Age: A continuous variable for the age of each family member in years.
10. Education: A continuous variable for the educational attainment of each family member
11. Occupation: A category variable for the occupation of each family member, recorded as (i) Share-cropping (ii) Cultivation of own land with hired labour (iii) Agricultural labour (iv) combination of share-cropping and agricultural labour (v) Services (Mason, Driver, Mechanic, Pastor, Bicycle Mechanic, Cook, Teacher) (vi) Factory Worker (vii) Vendor (viii) Commercial Enterprise (ix) Transfer from Family Member and (x) Professional Services (Lawyer, Nurse)
12. Religion: A category variable recorded as (i) Catholic (ii) Protestant (iii) Voodoo (iv) no reported religion.
13. Family Assets: Continuous variables for the number of assets such as cows, goats, chickens, pigs, donkeys, sheep, horses, ducks, cars, TVs, radios and luxury appliances.
14. Behavioural Characteristics: Category variables for information on use of soap, diapers and hand-washing habits.
15. Use of other water treatment systems: a category variable recorded as (i) Add chlorine sometimes (ii) Borrow filtered water from friends when sick (iii) Boil water when sick (v) Boil water always.
16. Filter use characteristics: Continuous variables for the date of last cleaning, year installed and rate of use.
17. Reason no filter: A category variable for the reason families did not own filters.
18. Household hygiene: An observational measure of hygiene on a scale from 1 to 5.

4.6 Composite Indicators

Inclusion of the entire vector of survey variables in a statistical analysis procedure such as a regression would create problems with multicollinearity, owing to the inter-related nature of the variables, and result in statistically insignificant results given the parsimonious sample size. Nor would inclusion of individual variables such as the number of donkeys owned by a family be particularly meaningful or helpful in explaining differential health impacts. It was therefore necessary to collapse the extensive vector of survey variables into composite variables that compressed relevant information into a more useful form.

1. Composite Wealth Variable: Family assets were multiplied with average asset values to generate a cumulative asset value of observable family assets for each family. However, this cumulative variable is incomplete as a measure of family wealth as it omits assets such as bank accounts and expatriate incomes from family members in the United States. I therefore decided to combine this variable with the quality of housing variable to produce a categorical measure of relative wealth ranging from 1 to 5, weighting quality of housing by 75% and the cumulative asset value variable by 25%. I accorded a higher weight to the quality of housing variable, as health outcomes are more likely to be influenced by investments in housing than in the other non-visible assets that I might have missed in the cumulative asset value variable.
2. Composite Hygiene Variable: The observational hygiene variable was combined with the see-soap variable, which checked whether the household had soap. A higher weight was placed on the observational hygiene variable since the absence of soap might have indicated a temporary unavailability.
3. Family Educational Deficit: In considering the impact of educational attainment on health outcomes, I thought it appropriate to consider the average family education deficit as a relevant explanatory variable rather than an individual's education level. This captures the impact of parental educational levels on family health outcomes. Each individual's deficit was worked out as the difference between the educational level the individual should have attained for her age with no breaks in the education process and the individual's attained level. The optimal level for an adult was defined as a university degree.
4. Adult Education Deficit: Since children contribute little to the educational deficit, families with high numbers of children register a low average family educational deficit. The adult education deficit improves on the average family measure by considering only the average of the education deficit of persons above 15 years of age. This would include the members of the family mainly involved in household tasks such as cooking, cleaning and looking after children and other tasks in which

educational attainment can be expected to improve hygiene practices relevant to health outcomes.

4.7 Statistical Methods

4.7.1 Univariate Analysis

Simple summations were used to provide a preliminary description of the data. These were used to describe variables of interest in the population such as religious affiliation, quality of housing, quality of sanitation, educational attainment, compliance with filter use rules, etc. These descriptive statistics were useful in understanding the demographic features and behavioural characteristics of the population.

4.7.2 Bivariate Analysis

As a precursor to the development of multivariate models, a number of bivariate analyses were performed to detect trends in the data. These analyses sought to establish whether there were differences in the mean values of particular variables between groups of interest in the population. For instance, a bivariate analysis was used to compare the incidence of diarrhea in groups with and without filters, not controlling for any other variables. Bivariate analyses were also used to examine differences in population characteristics, such as quality of housing, assets and sanitation facilities, between the intervention and non-intervention groups.

Particular statistical tests are available to establish whether the difference in mean levels of a variable between two populations is in fact significant at a specified level of confidence. Analytic methods used include t-tests, comparison of differences in proportions, Pearson's goodness-of-fit tests, Fisher tests and analysis of variance.

4.7.2.1 Assessing The Difference Of Two Averages Using Independent T-Tests

As an illustration of t-tests, consider two populations, one of which is subject to a particular treatment such as access to water treatment filters. We are interested in comparing the mean of a particular outcome, say the number of incidences of diarrhea in a month, in the treatment group with the mean of the same outcome in the control group, which does not use filters. Independent t-tests facilitate such a comparison.

An independent t-test computes the relevant means in each group and assesses whether the difference of the means is zero at a given level of statistical confidence. The test first computes the difference in means from the available samples. It then computes the variance of the mean in each group and combines the two variances using a pooled variance estimator if the two population variances can be considered equal. If the variances are of different magnitude, data transformations may be necessary to achieve uniform variance (Berthouex, 1994). The variance of the difference in means is then computed. This procedure is robust to moderate non-normality because the central limit effect will tend to make the distributions of the averages and their difference normal even though the parent distributions are not normal. It is now possible to generate a distribution of the difference of means using a Student-t distribution. If the value 0 lies in the specified confidence interval, we conclude that the observed difference in means is not statistically significant at the specified confidence level.

4.7.2.2 Assessing The Difference Of Proportions Using The Normal Approximation To The Binomial Distribution

Ratios and proportions are common in biological and epidemiological studies. In the health impact study documented in this thesis, the incidence of diarrhea has been recorded as a dichotomous variable. The average of this variable across the population of interest is in fact the proportion of that population that has suffered an incidence of diarrhea in the past month.

Ideally, a dichotomous variable ought to be modeled using a binomial probability distribution. However, the binomial distribution in inconvenient to use with a large sample size. Under certain circumstances the normal distribution provides a good approximation to the binomial distribution. The binomial distribution is symmetric when p, the mean incidence of the event, is equal to 0.5 (Berthouex, 1994). For values of $p<0.5$, it is skewed to the right; for values of $p>0.5$ it is skewed to the left. For a large sample size n, however, the skewness is not great unless p is near 0 or 1. The distribution approaches symmetry as n becomes larger, the approach being more rapid when p is close to 0.5. More importantly to this study, the binomial distribution can be approximated by a normal distribution with the same mean and variance. This approximation gives reasonable results if $np>5$ *and* $n(1-p)>5$. Both these conditions are met in this study. Knowing the difference in proportion in each population, one needs to determine the variance of this difference using a suitable approximation. Thereafter the distribution of the difference in proportions can be plotted as a normal distribution.

4.7.2.3 Assessing Goodness-Of-Fit With Pearson's Chi-Square Test

In general, any procedure that seeks to determine whether a set of data could reasonably have originated from some given probability distribution is called a goodness-of-fit test (Larsen, 1981). The principle behind this test is to first group the observed variable according to category variables of interest. Then each category's expected occupancy is calculated on the basis of the presumed model. If the set of expected and observed frequencies are inconsistent, the conclusion would be that the model was inappropriate. The chi-square distribution is used to assess the inconsistency between expected and observed frequencies at a specified level of statistical confidence.

The version of the goodness-of-fit test used here is to examine the hypothesis of independence between variables. For instance, we may want to test whether incidence of diarrhea is independent of distribution of filter. Data for a test of independence are presented in tabular form with rows representing the categories of one criteria and columns the categories of the other. Such displays are often called contingency tables.

This reliable and uncomplicated procedure is the mainstay of the bivariate analyses used in this study.

4.7.2.4 Comparing Means Using Analysis Of Variance

Analysis of variance, ANOVA, tests two or more treatment groups to determine whether their means could have been obtained from populations with the same true mean. This is done by estimating the amount of variation within treatments and comparing it with the variance between treatments. If the treatments are from populations with the same mean, as far as can be determined from evidence of the available data, the variation within each variation within each treatment will be about the same as the variation between treatments (Berthouex, 1994). The F statistic is used to test whether the within variance and the between variance are alike. The F statistic measures variability in estimated variances in the same manner that the t statistic is a measure of the variability in estimates of means. ANOVA is a particularly attractive alternative to multiple t-tests that should be conducted to compare more than two means.

4.7.3 Multivariate Analysis

It is often not possible to randomize the assignment of the intervention to be tested, as explained in 3.3.1. It is even less possible to create treatment and control populations that are identical in all respects other than their access to the intervention. Wonnacott and Wonnacott (1987) stress the importance of multivariate analysis in observational studies. "Granted the need for observational studies, and at the same time their bias due to uncontrolled extraneous factors, how can we make them as careful and valid as possible? The problem is, we no longer can make the treatment and control groups equal by randomizing. How then can we make them equal? We record and observe the extraneous factors. Then, instead of designing them constant (as we would ideally do), we analyze our data in a compensating way (that gives us, insofar as possible, the same answer as if we could have held them constant.) This analysis of the data is precisely what regression provides."

In the context of this thesis, the problem is therefore to create a model that helps explain health outcomes in terms of relevant explanatory variables. In general, an equation being fitted to data may be building a purely descriptive empirical model or a mechanistic model based on a fundamental picture of the process being modeled. In this case, the model being generated is expected to provide a descriptive understanding of health outcomes -- there is no fundamental picture of the determinants of health outcomes to which our data or model must adhere. A response variable has been measured at several settings of the independent variables. Regression is the process of fitting an equation to this data.

The models used in this analysis include ordinary least squares, logistic and probabilistic regression models. While ordinary least squares regression assumes a linear model, logistic and probabilistic regression use a non-linear approach. The term linear refers to the parameters in the model and not to the independent variables. In regression, numerical values are known for the dependent and independent variables once the experiment or survey has been completed. It is the parameters that are unknown and must be computed. In a linear model, these parameters can be written as a linear function of the dependent variable. The independent variables themselves can be non-linear. For instance, in this analysis the incidence of diarrhea showed a quadratic relationship with age. This necessitated the inclusion of a squared-term for age. This did not affect the linear nature of the model.

In a non-linear model, parameters cannot be determined as a linear combination of the dependent variables. For instance, the cumulative distribution function in a logistic model involves exponential terms. The parameters are estimated using numerical methods.

All three techniques qualify as parametric estimations, which imply the assumption of a specific underlying probability distribution function in estimating the model. The ordinary least square model assumes, for instance, that the error term is normally distributed.

4.7.3.1 Ordinary Least Squares Regression

The method of ordinary least square regression computes model parameter values by minimizing the sum of squared errors terms or residuals. The error term or residual is the difference between the actual value of the response variable and its predicted value in the model. If the residuals are normally and independently distributed with constant variance, the parameter estimates are unbiased and have minimum variance. More formally the assumptions of the classical linear regression model are:

(i) The relationship between the response variable and the explanatory variable parameters is linear, as defined above.

(ii) The explanatory variables are non-stochastic variables.

(iii) The error terms shows a constant variance regardless of the value of the explanatory variables. The mean of the error at each value of the explanatory variable is zero.

(iv) The error terms are random and are not correlated with each other. The error term may be regarded as the sum of two components. (a) Measurement error, the result of reporting, recording or measurement inaccuracies. (b) Stochastic error, the result of inherent irreproducibility of biological and social phenomena (Wonnacott, 1987). This type of error may be reduced by controlling for all extraneous factors, as has been attempted in this survey.

The Gauss Markov theorem states that if the above conditions are met, parameter estimates computed by the method of least squares will be the best linear unbiased estimators of the true model parameters. In other words, of the whole universe of unbiased linear estimators the coefficient estimates from the ordinary least squares method has the lowest variance. Unbiased in this context means that the expected value of the parameter estimates in repeated samples would equal the true parameter values.

For models that are linear in the parameters, there is a simple algebraic solution for the least square parameter estimates. As the number of parameters increases, algebraic solutions are still available but are harder to compute. These calculations are performed using matrix algebra. No unique algebraic solution exists for non-linear models; parameter estimates are determined by iterative numerical methods (Berthouex, 1994).

Having obtained the mean value of the parameters, it is now necessary to estimate the statistical precision of the estimate. If the Gauss Markov assumptions are met, the parameters will be normally distributed over the mean value. The ordinary least squares minimisation process will generate algebraic equations for the variance of each parameter. These will depend on, amongst other factors, the experimental error variance, which can be estimated indirectly from the residual sum of squares if the model is correctly specified. A t-distribution can now be used to generate a range of parameter estimates at a given level of statistical confidence.

4.7.3.2 The Linear Probability Model

The estimation of a binary choice model, with a dichotomous dependent variable, by ordinary least squares regression results in a linear probability model. The main problem with the linear probability model is that the specification cannot be truly linear since the residuals are heteroskedastic, or show non-constant error variance. However, it is possible to apply least squares regression to a dichotomous dependent variable as if the dependent variable was a continuous, unbounded interval variable. A second problem that arises is with predictions that lie outside the 0-1 bounds. The main advantage of linear probability models is that one can interpret the coefficient estimates as effects of changes in the independent variables on the probability of the response variable outcome.

4.7.3.3 Probit Regression

A more involved way of dealing with binary response data is to use non-linear probability models such as probit and logit (discussed below). The idea behind both models is the same: suppose that the binary response variable D comes in fact from an unobserved latent variable Y such that when Y is bigger than some threshold (usually 0) we observe an answer D=1 and when it is lower we observe an answer D=0. Now suppose that Y is in fact linearly determined by our predictors. Then changing the predictors will change Y and thus also the observed probability of getting a D=1. The probit model is based on this idea plus the assumption that Y is normally distributed with mean 0 and standard deviation 1. The logit model assumes that Y has a logistic distribution (something close to the normal but with slightly fatter tails). Using these assumptions, the model is then estimated using maximum likelihood estimation.

It should be noted however that the estimated coefficients using probit analysis refer to the effect of the predictor on the underlying unobserved variable Y. Statistical programming packages facilitate the conversion of probit estimated coefficients to the more useful and intuitive linear probability framework.

4.7.3.4 Logistic Regression

As mentioned above, the logit model is a latent variable model in which the observed variable D becomes 1 if the underlying Y is greater than zero and becomes zero otherwise. One of the problems with the linear probability model was that the estimated probability function needed to be within 0 and 1. Now suppose that instead of estimating the probability p as a linear function of the predictors, we assume that the odds in favour of a positive outcome, p/(1-p), are in fact a linear function of the predictors. The odds ratio p/(1-p) can now take any value between 0 and infinity, so the model is less constrained than before. However, the model may yet generate a negative predicted odds ratio. To avoid this final obstacle, the model assumes that the log of the odds ratio ln p/(1-p) is a linear function of the predictors. Since the log takes any value between minus and plus infinity, the model is unconstrained. The above transformation is called a logistic transformation and the models using it are called logistic models.

4.7.4 Violations of the Gauss Markov Assumptions

It is improbable that a sample will satisfy all the Gauss Markov assumptions. This section lists common violations of these assumptions, sources of the problem, their impact of OLS estimates and possible fixes.

4.7.4.1 Heteroskedasticity: This refers to the existence of non-constant error variance. In other words, the variance of the errors depends on the observation. This might occur if the units in the sample are of varying size or if the dependent variable is an average such that the sample underlying that average varies across observation in the sample. It is guaranteed to happen in OLS regressions of a binary variable in which the error terms can assume only two values. Heteroskedasticity results in biased and inefficient standard errors, which reduces the accuracy of predicting model parameter estimates. Parameter estimates remain unbiased, however. The problem may be corrected by a weighted least squares correction that puts less weight on observations with a large error variance. Another technique is to calculate robust standard errors and settle for inefficient estimates.

4.7.4.2 Omitted Variable Bias: If the model leaves out variables that affect the response variable and are also correlated to the explanatory variables, there will be correlation in the error terms. This will bias the coefficients on the explanatory variables which is the most serious of all model estimation problems. Possible solutions are to collect information on all omitted factors and include these in the model. Another strategy is to look for an experiment in which the treatment effect is randomly assigned. A third approach is to seek instrumental variables that are correlated with the variables of interest but not correlated with the error term.

4.7.4.3 Selection Bias: It may happen that in order for an observation to get into the sample, its response variable must exceed a certain threshold. This causes selection bias in which the explanatory variables and the error term are correlated. The impact on parameter estimates is similar to that in omitted variable bias. Possible solutions are to evaluate the effect of the sample selection criterion or to use an instrumental variable to model the behaviour that determines whether an observation/individual is in the sample.

4.7.4.4 Measurement Error: In many situations, including the one investigated in this thesis, it is only possible to measure the dependent variable with error. If the measurement error is stochastically independent of the dependent variable and the explanatory variables, and has zero mean, then measurement error in a dependent variable does not cause biased estimates.

Measurement error in the explanatory variable will cause estimates to be biased towards 0, which is also know as attenuation bias. Solutions include instrumental variables and finding estimates of the variance of the measurement error.

4.7.4.5 Multicollinearity: Multicollinearity occurs if the independent variables in a regression analysis are highly correlated. In such a situation, standard errors will be large, although parameter estimates will be unbiased. Solutions include getting more data, combining the highly correlated variables and testing the joint effect of the variables. Multicollinearity has been a problem with the analysis of data in this survey because it restricted the number of interaction variables that could be created. It also raised standard errors and lowered the precision of parameter estimates in general, since many of the independent variables were substantially correlated.

4.8 Results

All statistical analyses were conducted in *Stata* statistical analysis software.

4.8.1 Univariate Analysis

4.8.1.1 Sample Size: The intervention group, which owned a GWI filter, comprised 62 households with a total of 461 individuals. The non-intervention group, which did not own a GWI filter, comprised 58 households with a total of 380 individuals.

4.8.1.2 Sex: The sample comprised 394 males and 447 females.

4.8.1.3 Age: The sample comprised 138 children less than or equal to five years of age; 221 children older than five but younger than 16; and 482 adults 16 years of age and older.

4.8.1.4 Diarrheal incidence: 86 persons, or 10.23%, of the sample population of 841 persons, reported an episode of diarrhea in the month before the survey.

4.8.1.5 Educational Deficit: The average educational deficit of the 120 families in the sample was 6.4 years, with a standard deviation of 2.7 years. The sample showed a mean adult educational deficit, defined as the average education deficit of family members above age 15, of 10.2 years with a standard deviation of 3.0 years.

4.8.1.6 Household quality: Of 120 households, 35 (29.2%) lived in houses classified as quality rank 1. 40 (33.3%) families lived in houses of quality rank 1.5. The number of families that lived in houses of quality rank 2 and 2.5 were 21 (17.5%) and 18 (15%). Only 6 families (5%) lived in houses of quality rank 3.

4.8.1.7 Sanitation Facilities: 58.9% of the 120 families used a private bathroom. 20.6% of the families used a common or community bathroom. 20.5% did not have access to any sanitation facilities.

4.8.1.8 Source of Water: 51 families (43%) used piped spring water from a community tap. 52 families (44%) drew water from common groundwater wells operated by a hand pump. 10 families (8.5%) used water from both the above sources. 4 families used unpiped spring water. 1 family used more than two sources. No data was available for two families.

4.8.1.9 Electricity: 61.7% of families surveyed had no electricity. 18.3% had a legal connection. 18.3% had an illegal connection. 2 families had a generator and an illegal connection.

4.8.1.10 Assets: 72% of sample households had observable assets (including farm animals, household appliances and vehicles) estimated at under Haitian $2,000 (US$400). 14% had visible assets estimated at between $Haitian 2,000 and $4,000 (US$400 to US$ 800).

4.8.1.11 Work: 160 persons of the 841-person sample population reported being engaged in work. 34 persons (21% of the working population) worked the land in a share-cropping arrangement. 31 persons (19%) worked as agricultural labourers. 11 persons (6.8%) worked both as share-croppers and agricultural labourers. 5 (3%) persons hired labour to work their own land. 20 persons (12.4%) provided services including masonry, driving, automobile and bicycle mechanical work, cooking, teaching and religious activities. 29 persons (18%) worked as vendors. The remaining working population performed some combination of these functions. The sample included one factory worker and two persons who owned commercial enterprises. Six persons (3.7%) were engaged in professional services as lawyers and nurses.

4.8.1.12 Hygiene: On a subjective observed hygiene index of 1-5, 16 families (13%) recorded a score less than 2. Another 34 (28%) families registered a score greater than or equal to 2 and less than 3. 45 (37 %) families had a score greater than or equal to 3 and less than 4. As many as 25 (20%) families had a score of 4 and above.

4.8.1.13 Soap: 109 families (90%) said they used soap. Only 76% of the sample had soap in the house at the time of the survey.

4.8.1.14 Use of soap: Of 103 families on which data on use of soap is available, 45% said soap was used for bathing. 17% said soap was used for bathing and hand-washing. 4% said soap was used for bathing, hand washing and clothes-washing. 25% said soap was used for bathing and washing clothes.

4.8.1.15 Hand-washing: 100% of the 120 respondents said they washed their hands. 5% said they washed their hands before cooking. 22.5% said they washed their hands before eating and after toilet. 28.3% said they washed their hands after toilet. 11% said they washed their hands after toilet and after work. 12% said they washed their hands after work. Another 6% said they washed their hands before eating and after work.

4.8.1.16 Diapers: Of 41 families with children under age 2, 22 (53%) did not report use of diapers. 17 families used cloth diapers and two reported use of disposable diapers.

4.8.1.17 Water Treatment: Of 58 families that did not use the filter, 72.5% never treated their water in any way. 11 families (19%) used chlorine sometimes. 3 families (5%) borrowed filtered water from a GWI program member when a family member was sick. One family boiled water when a family member was sick. One family reported boiling water all the time. Of 62 families with filters, 92% never used any other type of treatment. One family sometimes used only chlorine and not the filter. Several families said they drank bottled water (brand name Culligan) when they went out (this is not considered as an alternative treatment here).

4.8.1.18 Rate of filter use: Of the 62 families with filters, 25% used their filter less than 0.5 times a day. About 45% used the filter between 0.5 and 1 time a day. Another 25% used the filter between 1 and 2 times a day. Only 4% used more than twice a day.

4.8.1.19 Reason no filter: Of 58 families without a filter, 6 families (9%) had no knowledge of the program. 25 families (43%) reported having no time to obtain the filter from the program office. 2 families (3.5%) said they could not afford the filter. 13 families (22.4%) said they had applied and were waiting for a filter. 3 families (5%) said they had missed obtaining the filter when it was being distributed. One family was confident that the source water was fine and that the filter was unnecessary. 8 families were sampled in an area not covered by the program.

4.8.2 Bivariate Analysis

4.8.2.1 Diarrheal incidence and related factors

4.8.2.1.1 Diarrheal incidence and GWI Filter:

- Contingency tables showed that 14.74% of the sample population without filters suffered a diarrheal episode compared with 6.51% of the sample population with filters. A Pearson chi square test of independence confirmed that the incidence of diarrhea was correlated with household filter ownership at 99% confidence.
- A Fisher test of association confirmed correlation between the two variables at 99% confidence.
- A t-test comparison of the mean value of diarrheal incidence in the intervention and non-intervention populations with Welch's correction for unequal variances showed that the mean incidence of diarrhea was higher in the non-intervention population. The result was significant at 99% confidence.
- A two-sample test of proportions also revealed that the proportion of persons affected by diarrhea in the intervention group was lower than that in the non-intervention population at 99% confidence.

4.8.2.1.2 Diarrheal incidence and GWI Filter across different age groups:

- In the under-6 age group, Pearson chi square tests and Fisher tests confirmed that the incidence of diarrhea was correlated with household filter ownership at 99% confidence.
- In the 6-15 age group, Pearson chi square tests and Fisher tests confirmed that the incidence of diarrhea was correlated with household filter ownership at 99% confidence.
- In the above-15 age group, Pearson chi square tests and Fisher tests confirmed that the incidence of diarrhea was correlated with household filter ownership at 95% and 94% confidence respectively.

4.8.2.1.3 Diarrheal incidence and Quality of Housing

- Contingency tables showed that 15.7% of families who lived in houses of quality rank 1 suffered an episode of diarrhea compared to 8.8% of families in houses of quality rank 1.5; 10.8% of families in houses of quality rank 2; 2.9% of families in houses of quality rank 2.5; and 9.8% of families in houses of quality rank 3. A Pearson chi square test showed that the two variables were correlated at 99% confidence.
- A t-test showed the population that did not suffer diarrhea had a higher quality rank of housing than the population that had experienced diarrhea. The result was significant at 99% confidence.

4.8.2.1.4 Diarrheal incidence and Sex

- Contingency tables showed that 9.4% of males suffered a diarrheal episode compared to 10.96% of women. A Pearson chi square test showed that the variables were clearly not associated at 95% confidence.
- A t-test showed that the male population suffered a slightly lower incidence of diarrheal disease but this result was only significant at 45% confidence.

- A two-sample test of proportions also revealed that the proportion affected by diarrhea in the male and female populations was different only at 45% confidence.

4.8.2.1.5 Diarrheal incidence and Age

- A contingency table showed that 31.16% of children under the age of 6 suffered a diarrheal episode, compared to 3.62% of children between age 6 and 16 and 7.62% of adults of 16 years and above. A Pearson chi square test confirmed that the variables were correlated at 99% confidence.
- An independent t-test showed that the mean age of the populations that had experienced a diarrheal episode was lower than the population that had not. This result was significant at 99% confidence.

4.8.2.1.6 Diarrheal incidence and Sanitation Facilities

- Contingency tables showed that only 7% of persons with a private bathroom experienced diarrhea, compared with 13% of the population who used a common bathroom and 16% of the population with no sanitation facilities. A Pearson chi-square test confirmed that the variables were associated with 99% confidence.
- An independent t-test showed the population that had experienced diarrhea had poorer sanitation facilities than the population free of diarrhea. The result was significant at 99% confidence.

4.8.2.1.7 Diarrheal incidence and Household Size

- A Pearson chi square test showed that mean family incidence of diarrhea and household size was not independent at 99% confidence.
- One-way analysis of variance suggested however that the correlation was not significant.

4.8.2.1.8 Diarrheal incidence and per capita filter water consumption

- In the intervention group, mean family incidence of diarrhea was found to be correlated to reported per capita consumption of filtered water. Diarrheal incidence tended to be lower at higher levels of water consumption. One way analysis of variance suggested that the correlation was significant at 98% confidence.

4.8.2.1.9 Diarrheal incidence and education

- Mean family incidence of diarrhea was found to be correlated with average family adult education at 95% confidence in a Pearson chi square test.
- An independent t-test showed that the average family adult education deficit of the population that had experienced a diarrheal episode was 10.5 years, compared to 9.8 years in the remaining population. This difference was significant at 93% significance. The sign or direction of this effect is the opposite of what is intuitively expected.

4.8.2.1.10 Diarrheal incidence and source of water

- Contingency tables showed that 11.4 % of the population that used piped spring water suffered a diarrheal episode, compared to only 7.23% of the population that used well water. A Pearson chi square test across all source water categories indicated that the variables were correlated at 90% confidence.

4.8.2.1.11 Diarrheal incidence and Total Coliform content

- The average family incidence of diarrhea was not independent of total coliform content of the stored household water at 99% confidence, according to a Pearson chi square test.
- An independent t-test showed that the mean household water total coliform

content in the population that had suffered diarrhea was higher than in the population unafflicted by diarrhea by 760 cfu. The difference was significant at 99% confidence. It must be remembered that the total coliform content of the water is based on a one-time measurement and the reported diarrhea variable covers a period of a month.

4.8.2.1.12 Diarrheal incidence and *E. coli* content

- The average family incidence of diarrhea was not independent of *E. coli* content of the stored household water at 99% confidence, according to a Pearson chi square test.
- An independent t-test showed that the mean household water *E. coli* content in the population that had suffered diarrhea was higher than in the population unafflicted by diarrhea by 108 cfu. The difference in means, however, was not statistically significant. It must be remembered that the *E. coli* content of the water is based on a one-time measurement and the reported diarrhea variable covers a period of a month.

4.8.2.1.13 Diarrheal incidence and Hygiene

- A Pearson chi square test showed that individual diarrheal incidence and the composite hygiene indicator (a weighted combination of the subjective observed hygiene variable and the presence of soap in the house) were correlated at 90% confidence.
- An independent t-test showed that the population that had experienced a diarrheal episode had a lower hygiene index than the remaining population. The result was significant at 99% confidence.

4.8.2.1.14 Diarrheal incidence and Assets

- A Pearson chi square test showed that individual diarrheal incidence and the composite asset indicator (a weighted combination of quality of housing and visible assets) were correlated at 94% confidence.

- An independent t-test showed that the population that had experienced a diarrheal episode had a lower level of assets than the remaining population. The result was significant at 99% confidence.

4.8.2.2 Filter Ownership and socio-economic factors

4.8.2.2.1 Filter Distribution and Quality of Housing:

- A Pearson chi-square test showed that the ownership of a filter and quality of housing were correlated at 94% confidence.
- An independent t-test showed that the population who owned a filter enjoyed a higher quality rank of housing that the remaining sample. This difference was small but statistically significant at 98% confidence.

4.8.2.2.2 Filter Ownership and Sanitation Facilities:

- Contingency tables showed that 63% of families with private bathrooms had a filter, compared to 46% of families who used a common bathroom and 27% of families who did not have any sanitation facilities. A Pearson chi-square test showed that the ownership of a filter and quality of sanitation facilities were associated at 99% confidence.
- An independent t-test showed that the population who owned a filter enjoyed a higher quality of sanitation facilities that the remaining sample. This difference was statistically significant at 99% confidence.

4.8.2.2.3 Filter Ownership and Asset Value:

- Ownership of a filter and asset values were not correlated, according to a Pearson chi-square test. I would attribute this result to the large number of asset value categories rather than an inherent lack of correlation between the variables. (This is probably not an effective test in the circumstances.)

- An independent t-test showed the population that owned a filter had a higher value of assets than the remaining sample. This difference was small but statistically significant at 98% confidence.

4.8.2.2.4 Filter Ownership and Adult Education Deficit

- An independent t-test showed that the population who owned a filter had a lower average adult education deficit (9.14 years) in their families than the non-intervention households (11.26 years). The difference was statistically significant at 99% confidence.

4.8.2.2.5 Filter Ownership and Household Size

- Contingency tables show that in households of less than 7 persons, the majority of families do not own filters. In households of between 7 and 12 persons, the majority of families own filters. However, a Pearson chi square test indicates that the two variables are not correlated at 95% confidence.
- An independent t-test showed that the population who owned a filter had a higher average household size (7.3 persons) in their families than the non-intervention households (6.3 persons). The difference was statistically significant at 93% confidence.

4.8.2.2.6 Filter Ownership and Age

- An independent t-test showed that the mean age of the population whose family owned a filter (22.9 years) was larger than the mean age of the population whose family did not own a filter (21.8 years). The difference was not statistically significant at 95% confidence.

4.8.2.2.7 Filter Ownership and Sex

- An independent t-test showed that the sex composition of the intervention and non-intervention population were not different at 95% confidence.

4.8.3 Multivariate Analysis

Using the trends evident in the bivariate analysis, this section employs multivariate regression models to describe the occurrence of diarrheal episodes in the sample population. Ordinary least squares regression is used as a first-pass estimation technique, followed by the more precise non-linear probit and logit regressions. The models are generated by including all variables found to be significant in the bivariate analysis and also their higher powers if appropriate. Variables that are found to be redundant to the model at an appropriate level of significance are abandoned on the basis of individual t-tests and F-tests of joint significance.

4.8.3.1 The Linear Probability Model

4.8.3.1.1 An OLS regression of individual occurrence of diarrhea against gwi (the filter variable), quality of housing, age and age2 (the square of age) revealed statistically significant coefficients on all variables at 95% confidence. Members of families owning filters were found to have a 6.7 percentage point lower probability of having experienced a diarrheal episode (in the month in question) than others of the same age and quality of housing rank. Improved quality of housing (which was treated as a continuous variable in this regression) was found to reduce the probability of housing. The probability of diarrhea showed a quadratic trend with age, with a minimum at age 37.

4.8.3.1.2: When a continuous variable for sanitation rank was included in the above OLS specification, the probability of filter households contracting diarrhea dropped to 5.65 percentage point lower than non-filter households, controlling for all other factors. This was consistent with the correlation earlier seen between filter ownership and quality of sanitation facilities on the one hand and the correlation between diarrheal incidence and quality of sanitation facilities on the other. The sign of the coefficient showed that the probability of contracting diarrhea decreased with improved sanitation facilities, controlling for all other factors. The effect of improved quality of housing continued to show the same trend as earlier but was not statistically significant at 95% confidence. This may be the result of collinearity between quality of housing and sanitation facilities, which causes standard errors to rise in a small sample. The age trend was identical to the earlier regression.

4.8.3.1.3: When the subjective observed hygiene indicator index is added to the model (again as a continuous variable), it indicated that the probability of contracting diarrhea decreased with improving hygiene. It was not statistically significant.

4.8.3.1.4: When the composite hygiene indicator is added to the model (as a continuous variable), it indicates that the probability of contracting diarrhea decreased with improving hygiene. It was statistically significant only at 75% confidence. In this specification, the magnitude of the effect of the filter diminishes to a 5.2 percentage point lower probability of contracting diarrhea, controlling for other factors. The impact of quality of housing falls further in magnitude and statistical significance, once again illustrating the impact of omitted variable bias when correlated variables are left out of the regression. Despite the correlation of explanatory variables, however, this specification showed a variable inflation factor of under 5, the ballpark figure for serious problems associated with multicollinearity.

4.8.3.1.5: When the composite category variable for asset values (the best available proxy for wealth) is included in the specification instead of the quality of housing variable, it shows much the same magnitude and significance as the variable it replaced. This is probably owing to the high weight given to quality of housing in determining relative asset values. (I shall use quality of housing as a variable that captures asset values hereafter, although the appendix contains several regressions with the composite asset variable).

4.8.3.1.6: When the family education deficit variable is added to the model, it shows a negative sign, suggesting that the incidence of diarrhea decreases with increasing family education deficit. This is the result of families with high numbers of children showing a low family education deficit, since the education deficit is least for children. Children experience more diarrhea than adults, so it follows that a low family education deficit may be associated with more diarrhea. The effect is not statically significant.

4.8.3.1.7: When the adult education deficit variable (which calculates the education deficit of only the adults in a family) is instead used, the sign continues to be negative. It is statistically significant at the 75% confidence level. The rest of the model looks much the same as earlier. This is the opposite effect of what would be expected.

4.8.3.1.8: The creation and inclusion of an interaction term between the filter variable and the quality of housing variable resulted in a high model standard error, which made most coefficients in the model statistically insignificant. Although this term would have had the potential to measure the differential impact of quality of housing in families with and without filters, the sample size is too small for this approach to be effective. In future regressions, effects on sub-populations of interest are examined by regressions specific to those populations. This is equivalent to creating an interaction term for each included variable with the specific variable of interest.

4.8.3.1.9: In a regression confined to children of age 5 and under, the filter reduced the probability of diarrhea by 16 percentage points, controlling for all other factors. The result was significant at 94% confidence. The other variables showed the same sign as in the larger model, but were not statistically significant. This may have been the result of multicollinearity in a small sample.

4.8.3.1.10: In a regression confined to children older than 5 years but under 16, the filter reduced the probability of diarrhea by 4 percentage points, controlling for all other factors. The result was significant at 85% confidence.

4.8.3.1.11: In a regression confined to adults of 16 years and older, the filter reduced the probability of diarrhea by 4 percentage points, controlling for all other factors. The result was significant at 85% confidence.

4.8.3.1.12: If the regression is confined to persons who draw water from piped springs (369 observations), the filter reduces incidence of diarrhea by 7.95 percentage points. This result is significant at 95% confidence. All other variables show the same trend as in the larger model.

4.8.3.1.13: If the regression is confined to persons who draw water from wells (332 observations), the filter reduces incidence of diarrhea by only 2.4 percentage points. This is not statistically significant. In this specification, the impact of quality of housing is greater than in the composite model.

4.8.3.1.14: When quality of housing and sanitation facilities are treated as category variables, the model shows much the same character as the original model. However, most variables lose statistical significance because the inclusion of several new dummy variables (4 for quality of housing, 2 for sanitation facilities) inflates the standard error. The sign of the regressions, however, are consistent with the model in which these quantities are treated as continuous variables. Since there is no real gain to the treatment of these quantities as category variables, they are therefore treated as continuous variables in this analysis. Care must be exercised in interpreting the coefficients on these variables. For instance, it must be recalled that moving from housing rank 1 to rank 2 does not imply doubling the standard of housing. It does however represent a change in ordinal rankings and therefore an increase in the quality of housing.

4.8.3.1.15: One of the problems of the OLS specification with a binary response variable is that it may generate values outside the 0-1 range which are meaningless to the linear probability framework. An analysis of the predicted values of the diarrhea variable in a representative regression showed that 92 out of 841 values were less than 0. This indicated that problems with the OLS specification were not unreasonable.

4.8.3.1.16: To correct for heteroskedasticity, or non-constant error variance, the weighted least squares and robust variance estimation techniques were used. Both techniques produced results only slightly different from the original OLS specification.

4.8.3.2 The Probit Model

The best fit model developed through OLS was applied to non-linear probit estimation. When the coefficients are reconverted to the linear probability framework, the model, which was corrected for heteroskedasticity using the robust variance estimation method, showed that:

- Individuals in filter-owning families had a 5.6 percentage point lower probability of contracting diarrhea (in the month preceding the survey) than the remaining sample population, controlling for other factors. This result was significant at 99% confidence.
- Individuals with access to improved sanitation facilities had a lower probability of contracting diarrhea. This result was significant at 94% confidence.
- Individuals who lived in higher quality houses had a lower probability of contracting diarrhea. This result was only significant at 75% confidence.
- Diarrheal incidence showed the same U-shaped quadratic trend as in the OLS model. Younger and older people were more likely to have diarrhea.
- Increasing levels of hygiene lessened the probability of diarrhea. This result was significant at 80% confidence.
- Increasing levels of average adult education deficit in a family reduced an individual's probability of having diarrhea. This result was significant at nearly 90% confidence.

It is thus apparent that the conclusions of the probit analysis are almost identical to the OLS framework.

4.8.3.3 The Logit Model

The best fit model developed through OLS was also applied to non-linear logit estimation. Coefficients have not been reconverted to the linear probability framework and are quoted as natural logarithms of the odds ratio. The model, which was corrected for heteroskedasticity using the robust variance estimation method, showed that:

sfodiarh	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
gwi	-.7098382	.243532	-2.91	0.004	-1.187152	-.2325243
qoh	-.3211465	.3128201	-1.03	0.305	-.9342627	.2919696
san	-.3434103	.1808985	-1.90	0.058	-.0111442	.6979648
age	-.1139203	.0226825	-5.02	0.000	-.1583772	-.0694633
age2	.0014449	.0003066	4.71	0.000	.0008441	.0020458
hyg	-.1209056	.0991954	-1.22	0.223	-.315325	.0735139
adeddef	-.0742053	.055377	-1.34	0.180	-.1827422	.0343316
_cons	.5020089	.8249453	0.61	0.543	-1.114854	2.118872

It is thus apparent that the logit analysis is similar in sign and significance to the OLS framework.

4.8.3.4 Mean Family Diarrheal Incidence Model

Apart from an analysis of the individual occurrence of diarrhea, I also constructed a model to explain mean family incidence of diarrhea. This model had the advantage of reducing collinearity between independent variables by considering each family as a discrete unit. (In the earlier models, all members of the same family shared a number of common characteristics. This prevented the independent variables from varying within a family and thus increased collinearity). The new model also has the advantage that the dependent variable is continuous and thus better suited to ordinary least squares regression analysis.

4.8.3.4.1 In a regression of famdiarh against filter-ownership, quality of housing, sanitation facilities, the composite hygiene variable and the adult education deficit variable, filter ownership reduced family incidence of diarrhea by 0.06 episodes per person per month. (The average family incidence of diarrhea amongst the 120 survey families was 0.11 episodes per person per month.) The result was significant at 95% confidence. Access to superior housing and sanitation, and better hygiene practices, all reduced the incidence of diarrhea. However, these results were not statistically significant at 95% confidence. Increasing adult education deficit also lowered the number of episodes of diarrhea.

4.8.3.4.2 When household size is added as a predictor to the model, it was not found to be statistically significant. Its sign implied that larger households experienced fewer episodes of diarrhea.

5. WATER QUALITY IMPACT STUDY

5.1 Objective

Previous studies undertaken for GWI used presence-or-absence testing to determine microbial contamination in community water sources and stored household water in Dumay. While a useful indicator of microbial contamination, presence-or-absence testing provides no information on the precise levels of contamination in source water. This was seen as a shortcoming in understanding the extent of the contamination problem and in the design of an optimal strategy to improve water quality.

This study sought to obtain precise quantitative estimates of the level of microbial contamination in different community water sources and stored household water by using the membrane filtration technique to measure levels of microbial indicator organisms such as total coliform and *Escherichia coli*.

5.2 Membrane Filtration

The membrane filtration method involves filtering a measured volume of sample, or an appropriate dilution of it, through a membrane filter that has a pore size of 0.45 microns and is usually made of cellulose esters (Hutton, 1983). A suitable volume of water, usually 100 ml for drinking water sources, is drawn through the filter by means of a vacuum or hand pump. Micro-organisms are retained on the filter surface which is then incubated face upwards on a suitable selective medium containing lactose. White absorbent pads are used to absorb the liquid media in the petridish. Sterile disposable plastic petri dishes with a nutrient pad already inserted and sterilized were used in this study. The medium used was m-ColiBlue24®, a lactose-based medium, containing inhibitors to selectively eliminate growth of non-coliforms. The total coliform colonies are highlighted by a non-selective dye, 2,3,5-Triphenoltetrazolium Chloride (TTC) which produces red-colour colonies. The *E. coli* colonies on the other hand, show as blue colonies, resulting from the action of a β-glucuronidase enzyme on 5-Bromo-4-Chloro-3-Indolyl-β-D-glucuronide (BCIG) (HACH 1999). The visible colonies are counted after an incubation period of 24 hours at 35±0.5ºC and expressed in terms of the number present in the volume of original sample.

Samples were collected in sterile Whirl-Pak bags and transported in a cooler fitted with an ice-pack. The samples were filtered within four hours of collection through a Millipore portable membrane filtration set-up. Filtration funnels were changed for each sample. The filter stand was sterilised using iso-propyl alcohol. The tweezers used to manipulate the member filter were sterilized in a candle flame.

The coliform group of organisms is used worldwide as indicators of faecal pollution, being normally associated with faeces and water. They are characterized broadly by their ability to ferment lactose at 35 or 37 C with the production of both acid, aldehyde and gas within 48 hours. However, it is an inconclusive indicator of faecal contamination as there are other sources of coliform organisms apart from the faeces of warm-blooded animals, namely vegetation and soil. Certain coliform organisms retain the ability to ferment lactose at 44C or 44.5C. These are faecal coliform organisms, which may sometimes be described as thermotolerant coliform organisms. These include the genus Escherichia and to a lesser extent occasional strains of Enterobacter, Citrobacter and Klebsiella. Of these organisms only *E. coli* is specifically of faecal origin being always present in the faeces of man, animals and birds in large numbers. It is rarely found in water or soil that has not been subject to faecal pollution. The presence of *E. coli* is regarded as strong evidence of faecal pollution. In temperate climates the faecal coliform organisms detected at 44C or 44.5 C consist predominantly of *E. coli* but in hot climates less than 50% of the faecal coliform organisms are probably *E. coli* (Hutton, 1983). The absence of faecal coliform organisms, especially when total coliform organisms have been detected does not necessarily rule out the possibility that faecal contamination may have occurred. In most cases, however, the absence of faecal coliform organisms in a water implies the water was not recently faecally contaminated. Therefore determinations of total coliform at 35 or 37C which result in the detection of a small number of coliform organisms (1-10 organisms per 100ml) may be of limited sanitary significance if faecal coliform organisms are absent also.

The medium used in this study facilitated the measurement of total coliform and *E. coli* colonies at 35 C. It did not facilitate measurement of the faecal coliform group of organisms. However, as described above, enumeration of *E. coli* colonies is a more precise indicator of faecal contamination than the faecal coliform group of organisms.

5.3 Sampling Methodology

A total of 25 community water sources were tested in duplicate. These included 22 wells and 3 piped spring captages. The sampling covered 16 out of 22 circuits.

Amongst the 58 families with no filters, tests of stored household water quality were conducted on 47 families. One family drank directly from source, two families did not have water stored in the house at the time of the survey and eight families' water could not be tested for logistic reasons (they were surveyed on the last day of my trip and processing takes 24 hours).

Amongst the 62 families with filters, tests of filtered water quality were performed on 34 families. Nine families did not have water in their filter at the time of the survey, one family had a dysfunctional filter, and no tests were performed on 17 families whose water had an adequate chlorine residual. (Tests were performed on a fraction of the families in whose water adequate chlorine residuals were detected and on all families in which chlorine residuals were low.)

5.4 Results

5.4.1 Source Water Quality Tests

The following table summarises the results of the source water quality tests:

Table 1: Source Water Quality Results

Circuit	Source	TC (1)	TC (2)	EC (1)	EC (2)
1A	S	240	240	1	2

Circuit	Source	TC (1)	TC (2)	EC (1)	EC (2)
3A	W	2	0	0	0
3B	W1	1	0	0	0
3B	W2	0	2	0	0
4B	W1	8	22	3	3
4B	W2	2	2	0	0
5A	W	0	0	0	0
5B	W	0	0	0	0
6A	S	232	240	9	9
6B	W1	3	0	0	0
6B	W2	7	4	0	0
6B	S	168	164	0	0
8A	W1	6	7	0	0
8A	W2	2	2	0	0
8B	W1	0	0	0	0
8B	W2	0	0	0	0
9B	W1	0	1	0	0
9B	W2	0	0	0	0
9B	W3	11	1	0	0
10A	W1	0	0	0	0
10A	W2	0	0	0	0
10B	W1	4	10	0	0
10B	W2	10	4	0	0
11A	W	0	2	1	2
11B	W	2	4	1	1

* TC = Total Coliform
* EC = *E. coli*
* S = Piped Spring Water
* W = Well Water

- The average total coliform count in well water over 23 wells was found to be 2.65 cfu/100ml with a standard deviation of 3.7.
- The average *E. coli* count in well water over 23 wells was found to be 0.28 cfu/100ml with a standard deviation of 0.7
- The average total coliform count in piped spring water over 3 spring taps was found to be 214 cfu/100ml with a standard deviation of 41.6.
- The average *E. coli* count in piped spring water over 3 spring taps was found to be 3.5 cfu/100ml with a standard deviation of 4.8.

Table: 2 Total Coliform in Source Water

Source	Number of Replicate Samples	Average Total Coliform Count (cfu/100ml)	Median Total Coliform Count (cfu/100ml)
Well Water	23	2.65	1
Piped Spring Water	3	214	236

Table: 3 *E. coli* in Source Water

Source	Number of Replicate Samples	Average *E. coli* Count (cfu/100ml)	Median *E. coli* Count (cfu/100ml)
Well Water	23	0.28	0
Piped Spring Water	3	3.5	1.5

The inability to sample more piped spring water taps in the village is a weakness in the study. Other MIT groups testing raw water at other taps found contamination of the same order of magnitude, however, suggesting that my results were representative.

5.4.2 Stored Water Quality Tests in Non-Intervention Households

- The average total coliform content in 42 households was 2,485 cfu/100ml, with a standard deviation of 1,870 and a median of 2,425.
- The average total coliform content in 19 households that drew piped spring water was 2,977 cfu/100ml, with a standard deviation of 1,879 and a median of 2,750.
- The average total coliform content in 21 households that drew water from wells was 2,213 cfu/100ml, with a standard deviation of 1,830 and a median of 2,250.
- An OLS regression showed that mean total coliform levels in stored water in non-intervention households decreased with better sanitation and better quality housing; it increased with adult education deficit and household size. None of the coefficients was statistically significant.
- The average *E. coli* content in 46 households was 162 cfu/100ml, with a standard deviation of 514 and a median of 6.
- The average *E. coli* content in 19 households that drew piped spring water was 232.5 cfu/100ml, with a standard deviation of 751 and a median of 5.
- The average *E. coli* content in 25 households that drew water from wells was 505 cfu/100ml, with a standard deviation of 1992 and a median of 5. This result was clearly influenced by an outlier.
- An OLS regression showed that mean *E. coli* levels in stored water in non-intervention households decreased with better sanitation; *E. coli* levels increased with adult education deficit, household size and better quality housing. The sanitation variable was significant at 93% confidence. None of the other coefficients was statistically significant.

Table 4: Total Coliform in Stored Water of Non-Intervention Households

	Number of Households Tested	Average Total Coliform Count (cfu/100ml)	Median Total Coliform Count (cfu/100ml)
Cumulative	42	2,485	2,425
Well Water Source	21	2,213	2,250
Piped Spring Water Source	19	2,977	2,750

Table 5: *E. coli* in Stored Water of Non-Intervention Households

	Number of Households Tested	Average *E. coli* Count (cfu/100ml)	Median *E. coli* Count (cfu/100ml)
Cumulative	46	376	6
Well Water Source	25	505	5.5
Piped Spring Water Source	19	232	5

5.4.3 Stored Water Quality Tests in Intervention Households

- In 11 filter-owning households with adequate residual levels of chlorine in both buckets, there was little evidence of microbial contamination, with average TC= 0.68 cfu/100ml (median TC=0 cfu/100ml) and average EC=0.05 cfu/100ml (median EC=0 cfu/100ml).
- In 10 filter-owning households with adequate residual levels of chlorine in the top bucket but inadequate chlorine in the bottom bucket, there was little evidence of microbial contamination, with average TC= 2.25 cfu/100ml (median TC=0.5 cfu/100ml) and average EC=0.1 cfu/100ml (median EC=0 cfu/100ml).
- In 1 filter-owning households with inadequate residual levels of chlorine in the top bucket but adequate chlorine in the bottom bucket, there was little evidence of

microbial contamination, with average TC= 10.5 cfu/100ml and average EC=0 cfu/100ml.

- In 10 filter-owning households with inadequate residual levels of chlorine in both buckets, there was substantial evidence of microbial contamination, with average TC=913 cfu/100ml (median TC=250 cfu/100ml) and average EC=22.3 cfu/100ml (median EC=0 cfu/100ml).

Table 6: Total Coliform and *E. coli* in Stored Water of Intervention Households

Population	Number	Average TC*	Median TC*	Average EC*	Median EC*
Intervention (Compliant)	11	0.68	0	0.05	0
Intervention (TB+,BB-)**	10	2.25	0.5	0.1	0
Intervention (TB-, BB+)	1	10.5	10.5	0	0
Intervention (TB-, BB-)	10	913	250	22.3	0

* units of cfu/100ml
** TB=Top Bucket, BB=Bottom Bucket
+ Correct level of chlorine
- Less than recommended level of chlorine

6. SUMMARY OF KEY RESULTS

6.1 Health Impact

- Of the total sample of 841 individuals, 86 (10.23%) were reported to have experienced at least one episode of diarrhea in the month preceding the survey.
- Of the 380 individuals with no access to a filter at home, 56 (14.74%) were reported to have experienced diarrhea in the month preceding the survey.
- Of the 461 individuals with access to a filter at home, only 30 (6.51%) were reported to have experienced diarrhea in the month preceding the survey.
- The population with access to a filter therefore experienced an 8.23 percentage point, or 56%, lower incidence of diarrhea than the population with filters. The result is statistically significant at p = 0.0001.
- Intervention and non-intervention populations could not be considered identical.
- Intervention households, who owned a filter, had a higher average quality of housing, better sanitation facilities, larger household size and lower education deficit than the non-intervention households.
- After controlling for divergences in socio-economic factors and observed hygiene, the filter was associated with a 5.2 percentage point lower probability of diarrhea. The result is statistically significant at 95% confidence.
- Improved sanitation facilities were independently associatied with lower diarrheal incidence at 95% confidence.
- Improved quality of housing was independently associatied with lower diarrheal incidence, but the result was not statistically significant at 95% confidence.
- Better hygiene was associated with lower diarrheal incidence, but the result was not statistically significant at 95% confidence.
- Age was closely correlated with diarrheal incidence, with younger and older persons more at risk. The result was statistically significant at 95% confidence.
- For children of age 5 and under, the filter was associated with a lower diarrheal probability of 16 percentage points, controlling for all other factors. The average incidence of diarrhea in the group was 31.16%.

- For older children in the 6-16 age group, the filter was associated with a lower diarrheal probability of 4 percentage points, controlling for all other factors. The result was not significant at 95% confidence.
- For persons of age 16 and older, the filter was associated with a 4 percentage point lower probability of diarrhea, controlling for all other factors. The result was not significant at 95% confidence.
- Of the 369 individuals who used piped spring water, 11.38% experienced a diarrheal episode.
- Of the 332 individuals who used well water, only 7.23% experienced a diarrheal epidose.
- Amongst the population that used piped spring water, the filter reduced the incidence of diarrhea by 7.7 percentage points.
- Amongst the population who used well water, the impact of the filter was not significant in explaining diarrheal incidence.

6.2 Water Quality Impact

6.2.1 Source Water Quality

- The average total coliform count in duplicate samples from 23 wells was 2.65 cfu/100 ml.
- The average *E. coli* count in duplicate samples from 23 wells was 0.28 cfu/100 ml.
- The average total coliform count in duplicate samples from 3 community taps that were supplied with piped spring water was 214 cfu/100 ml.
- The average *E. coli* count in duplicate samples from 3 community taps that were supplied with piped spring water was 3.5 cfu/100 ml.

6.2.2 Stored Water Quality in Intervention Households

- In 11 filter-owning households with adequate residual levels of chlorine in both buckets, there was little evidence of microbial contamination, with average TC= 0.68 cfu/100ml and average EC=0.05 cfu/100ml.

- In 10 filter-owning households with adequate residual levels of chlorine in the top bucket but inadequate chlorine in the bottom bucket, there was little evidence of microbial contamination, with average TC= 2.25 cfu/100ml and average EC=0.1 cfu/100ml.
- In 1 filter-owning households with inadequate residual levels of chlorine in the top bucket but adequate chlorine in the bottom bucket, there was little evidence of microbial contamination, with average TC= 10.5 cfu/100ml and average EC=0 cfu/100ml.
- In 10 filter-owning households with inadequate residual levels of chlorine in both buckets, there was substantial evidence of microbial contamination, with average TC=913 cfu/100ml and average EC=22.3 cfu/100ml.

6.2.3 Stored Water Quality in Non-Intervention Households

- In 42 non filter-owning families, there was evidence of considerable microbial contamination in stored water, with average TC = 2,484 cfu/100ml. In 45 non-filter owning famiiles, the average EC was found to be 162 cfu/100ml.
- Contamination levels were negatively associated with improved sanitation at statistically significant levels for both indicators.

7. DISCUSSION

The case study documented in this thesis investigated the impact of a chemical disinfectant-based point-of-use water treatment system in a rural area of Haiti. The study found that the intervention population experienced a lower incidence of diarrhea than the non-intervention population. Although part of this effect is attributable to better housing and sanitation facilities, the filter itself was independently associated with a 50% lower incidence of diarrhea at 95% confidence. The average incidence rate in the combined sample was approximately 10.5%. The greatest impact of the filter was felt in two populations: the under-5 age group, in which the filter was associated with a 16 percentage lower probability of diarrhea, and the population that used piped spring water. The impact of the filter on older populations was of a lower magnitude and was not statistically significant. The filter was not significantly associated with lower diarrheal incidence in the population that drew water from wells. Diarrheal incidence was not significantly correlated with an individual's sex, household size or adult educational attainment.

Source water quality analysis corroborated the findings of the health impact study. Water from wells was found to be largely free of faecal contamination. Piped spring water was slightly more contaminated in comparison. Importantly, stored water in non-intervention houses was considerably more contaminated than both types of source water, suggesting post-collection contamination is a major concern. Stored water in compliant intervention households was very pure.

Further investigation of water quality and diarrheal disease incidence trends at different times of the year may provide the basis for effective program restructuring. If well water is found to be of acceptable quality throughout the year (in terms of microbial contamination as well as potability indicators such as salinity and turbidity), GWI may consider encouraging the use of wells over piped spring water in Dumay. This may require a preliminary investigation of groundwater reserves followed by the development of new wells in areas currently served only by spring water. Access to pure water sources would need to be supplemented by interventions that promote the use of safe storage containers and behavioural change on the lines of the CDC's safe water system. In other words, it may no longer be necessary to promote GWI's two-bucket cotton-and-carbon filter system if access to clean source water can be extended to the whole community. Instead, the community could be provided with, or encouraged to buy, two special plastic buckets with a tap and a recessed opening that prevents the entry of hands and other large objects. The stored water would need to be dosed with a limited amount of disinfectant to prevent bacterial regrowth. This approach has the advantage of being cheaper and therefore more easily extended to the wider community. It would also reduce the community's exposure to potentially dangerous tri-halo methanes and other disinfection by-products, currently a concern with the filter.

This study examined only one village. It may not be possible to develop access to clean water sources in all the areas that GWI operates in, owing to financial constraints and/or unavailability of clean sources. GWI's current point-of-use filtration system is an appropriate intervention where clean water sources are unavailable. However, even in such communities, there is a strong case for extending safe storage interventions to the households that are yet to be provided with a filter.

The study also revealed that most diarrheal incidence occurred in the under-6 age group. This suggests that appropriate interventions aimed at families with children in this age group may be an effective means of reducing the overall incidence of diarrheal diseases.

Clearly, such program modifications would require considerable research into economic feasibility and social acceptability, as well as organizational and sustainability challenges. This study does not examine these issues; it merely draws attention to evidence that a point-of-use water treatment intervention based on safe storage, minimal disinfection and behavioural change may be a cost-effective means of increasing program impact in areas with access to relatively safe source water.

REFERENCES

Berthouex P., Brown L. 1994. *Statistics for Environmental Engineers.* CRC Publishers. Boca Raton.

HACH (1999) *HACH Method Specification Sheet For HACH Microbiological Methods – Total coliform and E. coli* HACH, USA.

Wateraid, 2002 "http://www.wateraid.org.uk/research/index.html"

World Bank 2002 "http://www.worldbank.org/data/countrydata/countrydata.html#AAG"

Hutton, L. 1983. *Field Testing of Water in Developing Countries.* Water Research Centre. Bucks, England.

Kimura A., Quick R., Thevos A., Tembo M., Shamputa I., Hutwagner L., Mintz E. 2001. Diarrhea Prevention through Household-Level Water Disnfection and Safe Storage in Zambia. *American Journal of Public Health.* 91 (10)

Larsen R., Marx, M. 1981. *An Introduction to Mathematical Statistics and its Applications.* Prentice Hall. New Jersey.

Makutsa P., Nzaku K., Ogutu P., Barasa P., Ombeki S., Mwaki A., Quick R. 2001. Challenges in Implementing a Point-of-Use Water Quality Intervention in Rural Kenya. *American Journal of Public Health.* 91 (10).

Mintz E., Bartram J., Lochery P., and Wegelin M. 2001. Not Just a Drop in the Bucket: Expanding Access to Point-of-Use Water Treatment Systems. *American Journal of Public Health.* 91 (10)

Mohamed F. 2001. *Household-level point-of-use water filtration in Haiti: Strategies for Program Management and Sustainability*

Mong Y., Kaiser R., Ibrahim D., Rasoatiana, Razamfimbololona L., Quick R.. 2001. Impact of the Safe Water System on Water Quality in Cyclone-Affected Communities in Madagascar. *American Journal of Public Health.* 91 (10).

Pan American Health Organization. 1998. *Health in the Americas.* Volume II.
Quick R., Venczel L., Gonzalez O., Mintz E., Highsmith A., Espada A., Ddamiani E., Bean N., De Hannover R., Tauze R. 1996. Narrow-mouthed water storage vessels and in-situ chlorination in a Bolivian community: a simple method to improve drinking water quality. *American Journal of Tropical Medicine and Hygiene*; 54: 511-516.

Rabe-Hesketh S., Everitt B. 2000. *A Handbook of Statistical Analysis Using Stata.* Chapman & Hall/CRC. Boca Raton.

Reiff F., Roses M., Venczel L., Quick R., Will V. 1996. Low Cost Safe Water for the

World: a Practical Interim Solution. *Health Policy*; 17: 389-408.

Reller M., Mong Y., Hoekstra R., Quick R. 2001. Cholera Prevention With Traditional and Novel Water Treatment Methods: An Outbreak Investigation in Fort-Dauphin, Madagascar. *American Journal of Public Health.* 91 (10)

Report WHO/UNICEF/WSSCC.

Roberts L., Chartier Y., Chartier G., Malenga G., Toole M., Rodka H. 2001. Keeping Clean Water Clean in a Malawi Refugee Camp: a Randomized Intervention Trial. *Bulletin of the World Health Organization.* 79: 280-287.

Semenza J., Roberts L., Henderson Aa., Bogan J., Rubin C. Water Distribution Systems and Diarrheal Disease Transmission: a case study in Uzbekistan. 1998. *American Journal of Tropical Medicine and Hygiene;* 59: 941-6.

Soleto L., Quick R., Venczel L., Mintz E., Aparicio J., Gironaz M., Hutwanger L., Greene K., Bopp C., Maloney K., Chavez Dd., Sobsey M., Tauxe R. 1999. Diarrhea Prevention in Bolivia through Point-of Use Disinfection and Safe Storage: a promising new strategy. *Epi Infect* ; 122: 83-90.

Stata Corporation. 1999. *Stata Reference Manual Release 6.* Stata Press. College Station, Texas.

The World Bank. 2002. *Country at a Glance Tables.*

United Nations Development Programme. 2001. *Human Development Report.*

Wateraid, Tearfund. 2001. *The Human Waste.*

WHO/UNICEF/WSSCC, 2000. Global Water Supply and Sanitation Assessment

Wonnacott T., Wonnacott R. 1987. *Regression: A Second Course in Statistics.* Robert E. Krieger Publishing Company. Malabar, Florida.

Appendix 1

SURVEY QUESTIONAIRE

FAMILY IDENTIFICATION NUMBER:

1. **How many members of your family live in your home?**

2. **What are the ages and gender of these members of your family?**

Num.	Relationship	Age	Sex

3. **In the past month, how often have these members of your family had the following illnesses:**

   (i) Diarrhoea (up to 5 loose stools on worst day)
   (ii) Heavy diarrhoea (up to 9 loose stools on worst day)
   (iii) Extremely heavy diarrhea (more than 10 loose stools on worst day)
   (iv) Fever

Num.	Diarrhea (up to 5 stools)	Heavy Diarrhea (5 to 9 stools)	Extreme Diarrhea (more than 10 stools)	Fever

4. **Where do you collect your drinking water from?**

Pond/Lake	Cistern	River	Well	Other

5. **Do you use a GWI filter?**

YES	NO (skip to 7)

6. **Since when have you used the filter?**

7. **Do you use any other kind of filter or water treatment device?**

YES	NO (skip to 13)

8. **What other water treatment system do you use?**

9. **How often do you use the GWI filter?**

10. **For which activities do you use filtered water?**

Drinking	Cooking	Bathing	Other

11. **Do you clean your filter?**

YES	NO (skip to 13)

12. **When did you last clean the filter?**

13. Do any members of the family drink unfiltered water outside the house?

Num.	Frequently	Occasionally	Rarely	Never

14. What sanitation facilities do your family use?

Private Bathroom	Common/Public Bathroom	Open	Other

15. Do you wash your hands?

YES	NO

16. When do you wash your hands?

Before Cooking	Before Eating	After Toilet	After Diaper Change	Other

17. Do you have soap?

YES	NO

18. Interviewer asks to see soap.

YES	NO

19. **When do you use soap?**

For Bath	For Hand-Washing	For Clothes Wash	Other

20. Do the babies wear diapers? (IF THERE ARE CHILDREN UNDER 2)

YES	NO

21. Are the diapers disposable or cloth?

DISPOSABLE	CLOTH

22. Do you wash your hands after changing the diapers?

YES	NO

23. How many members of your family work? (Internal household work

excluded).

Num.	Farm/Agriculture	Labour	Other (SPECIFY)

24. Please indicate who in this household has been to school and to which level they have studied.

Num.	Education Level

25. Observe or ask about the number of rooms in the house.

26. Note or ask about other household assets such as domestic animals, radio, et

cetera.

Radio	Cows	Goats	Chickens	Other

27. What religion do you follow?

28. Which church do you go to?

29. Test drinking water for total coliform and e-coli. One replicate.

	Sample	Replicate
Total Coliform		
E-Coli		

30. Test chlorine content in top and bottom bucket of filtration apparatus.

	Okay	High/Low (SPECIFY)
Top Bucket		
Bottom Bucket		

Appendix 2

Statistical Analysis

1. Description of Variables

variable name	storage type	display format	value label	variable label
familyid	str3	%9s		Family ID
circuit	str8	%9s		Circuit
hhs	byte	%8.0g		Household Size
sow	str5	%9s		Source of Water
gwi	byte	%8.0g		GWI Filter
yrinst	byte	%8.0g		Years Installed
rnf	str3	%9s		Reason No Filter
otrt	byte	%8.0g		Other Treatment
rou	float	%9.0g		Rate of Use
uia	str3	%9s		Used in Activities
dolc	byte	%8.0g		Date of Last Cleaning
sanit	str1	%9s		Sanitation
washhands	str5	%9s		Wash Hands
soap	str5	%9s		Soap
seesoap	byte	%8.0g		See Soap
useofsoap	str5	%9s		Use of Soap
diapers	byte	%8.0g		Diapers
disposable	byte	%8.0g		Disposable
rooms	str5	%9s		Rooms
qoh	float	%9.0g		Quality of Housing
electricity	str3	%9s		Electricity
hygiene	float	%9.0g		Hygiene
religion	str3	%9s		Religion
tv	byte	%8.0g		TV
radio	byte	%8.0g		Radio
cows	byte	%8.0g		Cows
goats	byte	%8.0g		Goats
chickens	byte	%8.0g		Chickens
pigs	byte	%8.0g		Pigs
donkeys	byte	%8.0g		Donkeys
sheep	byte	%8.0g		Sheep
horses	byte	%8.0g		Horses
ducks	byte	%8.0g		Ducks
car	byte	%8.0g		Car
luxapp	byte	%8.0g		Luxury Appliances
tb	str2	%9s		Top Bucket
bb	str2	%9s		Bottom Bucket
tc1	str4	%9s		Total Coliform
tc2	str4	%9s		Total Coliform (2)
ecoli1	str5	%9s		E-Coli
ecoli2	str4	%9s		E-Coli (2)
member	byte	%8.0g		Member
rship	str3	%9s		Relationship
age	float	%9.0g		Age
sex	byte	%10.0g		Sex
diarh	byte	%8.0g		Diarheaa
hdiarh	byte	%8.0g		Heavy Diarheaa
exdiarh	byte	%8.0g		Extreme Diarheaa
fever	byte	%8.0g		Fever
freq	byte	%8.0g		Frequently
ocss	byte	%8.0g		Ocassionally
never	byte	%8.0g		Never
ed	str2	%9s		Education
work	str5	%9s		Work
fmius	byte	%8.0g		Family Member in US
sfodiarh	float	%9.0g		Some Form of Diarheaa
eddef	byte	%8.0g		Education Deficit
hhs2	float	%9.0g		Household Size Squared
fameddef	float	%8.0g		Family Education Deficit

assval	float	%9.0g		Household Assets Value
tcav	float	%9.0g		Average Total Coliform
ecav	float	%9.0g		Average E-Coli
comp	float	%9.0g		Compliance
age2	float	%9.0g		Age Squared
san	long	%8.0g	san	Sanitation (numeric)
adeddef	float	%9.0g		Adult Education Deficit

2. Summary of Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
familyid	0				
circuit	0				
hhs	841	8.097503	2.908118	1	15
sow	0				
gwi	841	.548157	.4979717	0	1
yrinst	476	3.151261	1.257491	0	6
rnf	0				
otrt	841	.2853746	.8117516	0	5
rou	467	1.187773	.6412837	0	3
uia	0				
dolc	460	.9869565	.1135843	0	1
sanit	0				
washhands	0				
soap	0				
seesoap	841	.7764566	.4168673	0	1
useofsoap	0				
diapers	320	.5	.6133315	0	2
disposable	101	.3960396	.584452	0	2
rooms	0				
qoh	841	1.673603	.5957728	1	3
electricity	0				
hygiene	841	2.919738	.9244254	1	5
religion	0				
tv	841	.294887	.5173973	0	2
radio	841	1.034483	.8521681	0	4
cows	841	.7538644	1.278043	0	5
goats	841	.8703924	1.662211	0	10
chickens	841	2.98692	6.653039	0	50
pigs	841	.8834721	1.360401	0	8
donkeys	841	.156956	.5512775	0	3
sheep	841	.3460166	1.197373	0	7
horses	841	.020214	.1408155	0	1
ducks	841	.1474435	1.193094	0	11
car	841	.0071344	.0842134	0	1
luxapp	841	.0535077	.2251776	0	1
tb	0				
bb	0				
tc1	0				
tc1new	523	1434.465	1838.904	0	7000
tc2	0				
tc2new	480	1295.892	1657.924	0	6200
ecoli1	0				
ecoli1new	563	210.8099	1108.907	0	10000
ecoli2	0				
ecoli2new	530	90.71887	312.6794	0	2500
member	840	4.632143	2.880896	1	15
rship	0				
age	841	22.4247	17.89761	.1	100
sex	841	1.53151	.4993031	1	2
diarh	841	.0713436	.2575512	0	1
hdiarh	841	.0261593	.1597039	0	1
exdiarh	841	.0047562	.0688423	0	1
fever	841	.2021403	.4018353	0	1
freq	26	1	0	1	1
ocss	220	1	0	1	1
never	187	.9946524	.0731272	0	1
ed	0				

```
        work |         0
       fmius |        11           1           0           1           1
    sfodiarh |       841    .1022592    .3031692           0           1
       eddef |       841    6.047562    5.674724          -1          16
        hhs2 |       841    74.01665    49.94187           1         225
    fameddef |       841    6.047562    2.063743         1.4          16
      assval |       841     1880.82    2164.965           0       10650
        tcav |       523    1435.576     1771.75           0        6600
        ecav |       563    203.6936    1082.946           0       10000
        comp |       841    .2449465    .4303114           0           1
        age2 |       841    822.8104    1226.662         .01       10000
         san |       841    1.617122    .8052753           1           3
     adeddef |       841    9.899822     2.80884         3.1          16
```

2.1 Labels of Category Variables

```
electricity:
           0 None
           1 Regular
           2 Illegal
           3 Generator
diapers:
           0 None
           1 Yes
bb:
           0 Insufficient Chlorine
           1 Adequate
           2 Too High
tb:
           0 Insufficient Chlorine
           1 Adequate
           2 Too High
work:
           0 No work or source of income
           1 Share-Cropping
           2 Cultivation of own land using hired labour
           3 Agricultural Labour
           4 Share Cropping and Agricultural Labour
           5 Services
           6 Factory Worker
           7 Vendor
           8 Commercial Enterprise
           9 Transfer from family member
          10 Professional Services
religion:
           0 No Reported Religion
           1 Protestant
           2 Catholic
           3 Voodoo
useofsoap:
           1 Bath
           2 Hand Washing
           3 Clothes Wash
           4 When Dirty
washhands:
           1 Before Cooking
           2 Before Eating
           3 After Toilet
           4 After Diaper Change
           5 After WOrk
           6 When Dirty
san:
           1 No Facilities
           2 Common
           3 Private
dolc:
           1 Last Week
uia:
           1 Drinking
           2 Cooking
```

3 Washing Face

3. Univariate Data Description

3.1 Filter Distribution (Family as unit)

GWI Filter	Freq.	Percent	Cum.
0	58	48.33	48.33
1	62	51.67	100.00
Total	120	100.00	

3.2 Quality of Housing (Family as unit)

Quality of Housing	Freq.	Percent	Cum.
1	35	29.17	29.17
1.5	40	33.33	62.50
2	21	17.50	80.00
2.5	18	15.00	95.00
3	6	5.00	100.00
Total	120	100.00	

3.3 Sanitation Facilities (Family as unit)

Sanitation	Freq.	Percent	Cum.
3	68	56.67	56.67
2	26	21.67	78.33
1	26	21.67	100.00
Total	120	100.00	

3.4 Source of Water (Family as unit)

Source of Water	Freq.	Percent	Cum.
1	51	43.22	43.22
1&2	10	8.47	51.69
1&2&6	1	0.85	52.54
2	52	44.07	96.61
3	4	3.39	100.00
Total	118	100.00	

3.5 Household Size (Family as unit)

Household Size	Freq.	Percent	Cum.

1	3	2.50	2.50
2	6	5.00	7.50
3	5	4.17	11.67
4	13	10.83	22.50
5	18	15.00	37.50
6	11	9.17	46.67
7	14	11.67	58.33
8	15	12.50	70.83
9	12	10.00	80.83
10	11	9.17	90.00
11	4	3.33	93.33
12	3	2.50	95.83
13	2	1.67	97.50
14	2	1.67	99.17
15	1	0.83	100.00
Total	120	100.00	

3.6 Geographical Circuit (Family as unit)

```
tab circuit
```

Circuit	Freq.	Percent	Cum.
10A	2	1.67	1.67
10B	3	2.50	4.17
11A	2	1.67	5.83
11B	4	3.33	9.17
1A	6	5.00	14.17
1B	5	4.17	18.33
2A	4	3.33	21.67
2B	9	7.50	29.17
3A	9	7.50	36.67
3B	4	3.33	40.00
4A	5	4.17	44.17
4B	6	5.00	49.17
5A	9	7.50	56.67
5B	6	5.00	61.67
6A	6	5.00	66.67
6B	6	5.00	71.67
8A	9	7.50	79.17
8B	9	7.50	86.67
9A	4	3.33	90.00
9B	4	3.33	93.33
Beauge	3	2.50	95.83
Bonette	2	1.67	97.50
Bonnette	3	2.50	100.00
Total	120	100.00	

3.7 Per Capita Family Incidence of Diarrhea (Family as unit)

famdiarh	Freq.	Percent	Cum.
0	60	50.00	50.00
.0769231	2	1.67	51.67
.0833333	1	0.83	52.50
.0909091	3	2.50	55.00
.1	3	2.50	57.50
.1111111	3	2.50	60.00
.125	8	6.67	66.67
.1428571	7	5.83	72.50
.1666667	4	3.33	75.83
.2	3	2.50	78.33
.2222222	3	2.50	80.83
.25	7	5.83	86.67
.2727273	1	0.83	87.50
.2857143	3	2.50	90.00

.3333333	2	1.67	91.67
.4	2	1.67	93.33
.4285714	1	0.83	94.17
.5	5	4.17	98.33
.6	2	1.67	100.00
Total	120	100.00	

3.8 Per Capita Family Education Deficit (Family as unit)

Family Education Deficit	Freq.	Percent	Cum.
1.4	1	0.83	0.83
1.8	1	0.83	1.67
2	1	0.83	2.50
2.375	1	0.83	3.33
2.428571	1	0.83	4.17
2.666667	1	0.83	5.00
2.75	2	1.67	6.67
3.5	2	1.67	8.33
3.714286	2	1.67	10.00
3.75	1	0.83	10.83
3.777778	1	0.83	11.67
3.8	1	0.83	12.50
3.833333	1	0.83	13.33
3.875	1	0.83	14.17
4	4	3.33	17.50
4.1	1	0.83	18.33
4.142857	1	0.83	19.17
4.2	1	0.83	20.00
4.272727	1	0.83	20.83
4.4	1	0.83	21.67
4.428571	1	0.83	22.50
4.444445	1	0.83	23.33
4.5	1	0.83	24.17
4.666667	2	1.67	25.83
4.7	1	0.83	26.67
4.714286	1	0.83	27.50
4.75	2	1.67	29.17
4.833333	2	1.67	30.83
4.857143	1	0.83	31.67
5	1	0.83	32.50
5.125	1	0.83	33.33
5.2	1	0.83	34.17
5.230769	1	0.83	35.00
5.333333	1	0.83	35.83
5.384615	1	0.83	36.67
5.4	2	1.67	38.33
5.416667	1	0.83	39.17
5.444445	1	0.83	40.00
5.5	1	0.83	40.83
5.6	2	1.67	42.50
5.642857	1	0.83	43.33
5.769231	1	0.83	44.17
5.8	3	2.50	46.67
5.909091	1	0.83	47.50
6	4	3.33	50.83
6.090909	1	0.83	51.67
6.25	1	0.83	52.50
6.285714	1	0.83	53.33
6.375	1	0.83	54.17
6.4	1	0.83	55.00
6.454545	1	0.83	55.83
6.5	1	0.83	56.67
6.571429	1	0.83	57.50
6.625	1	0.83	58.33
6.666667	1	0.83	59.17

6.714286	1	0.83	60.00
6.75	1	0.83	60.83
6.8	1	0.83	61.67
6.875	1	0.83	62.50
6.888889	1	0.83	63.33
6.9	1	0.83	64.17
7	5	4.17	68.33
7.125	1	0.83	69.17
7.142857	1	0.83	70.00
7.214286	1	0.83	70.83
7.285714	1	0.83	71.67
7.333333	1	0.83	72.50
7.4	1	0.83	73.33
7.571429	1	0.83	74.17
7.625	1	0.83	75.00
7.714286	1	0.83	75.83
7.75	1	0.83	76.67
7.866667	1	0.83	77.50
8	3	2.50	80.00
8.1	1	0.83	80.83
8.181818	1	0.83	81.67
8.25	1	0.83	82.50
8.333333	1	0.83	83.33
8.5	2	1.67	85.00
8.6	1	0.83	85.83
8.666667	1	0.83	86.67
8.777778	1	0.83	87.50
9	1	0.83	88.33
9.2	1	0.83	89.17
9.5	1	0.83	90.00
9.75	1	0.83	90.83
9.8	1	0.83	91.67
10.14286	1	0.83	92.50
10.33333	1	0.83	93.33
11	1	0.83	94.17
12.5	1	0.83	95.00
13	2	1.67	96.67
13.25	1	0.83	97.50
14	1	0.83	98.33
16	2	1.67	100.00
Total	120	100.00	

3.9 Adult Education Deficit (Family as unit)

Adult Education Deficit	Freq.	Percent	Cum.
3.1	1	0.83	0.83
3.5	2	1.67	2.50
3.8	1	0.83	3.33
4.4	1	0.83	4.17
4.5	1	0.83	5.00
5.5	1	0.83	5.83
6.2	2	1.67	7.50
6.3	1	0.83	8.33
6.4	1	0.83	9.17
6.5	2	1.67	10.83
6.7	2	1.67	12.50
7	6	5.00	17.50
7.1	1	0.83	18.33
7.3	1	0.83	19.17
7.4	1	0.83	20.00
7.5	2	1.67	21.67
7.6	2	1.67	23.33
7.8	1	0.83	24.17
8	3	2.50	26.67
8.1	2	1.67	28.33

8.25	4	3.33	31.67
8.4	1	0.83	32.50
8.5	1	0.83	33.33
8.8	2	1.67	35.00
8.9	1	0.83	35.83
9	4	3.33	39.17
9.2	2	1.67	40.83
9.3	1	0.83	41.67
9.4	3	2.50	44.17
9.5	3	2.50	46.67
9.7	1	0.83	47.50
10	4	3.33	50.83
10.1	2	1.67	52.50
10.2	1	0.83	53.33
10.4	1	0.83	54.17
10.5	2	1.67	55.83
10.75	1	0.83	56.67
10.8	1	0.83	57.50
11	5	4.17	61.67
11.25	1	0.83	62.50
11.3	1	0.83	63.33
11.5	6	5.00	68.33
11.6	1	0.83	69.17
11.8	1	0.83	70.00
11.9	1	0.83	70.83
12.3	1	0.83	71.67
12.5	4	3.33	75.00
12.6	2	1.67	76.67
12.7	1	0.83	77.50
12.75	1	0.83	78.33
13	4	3.33	81.67
13.25	2	1.67	83.33
13.3	2	1.67	85.00
13.5	1	0.83	85.83
13.8	1	0.83	86.67
14	6	5.00	91.67
14.7	2	1.67	93.33
15.5	1	0.83	94.17
16	7	5.83	100.00
Total	120	100.00	

3.10 Household Asset Value (Family as unit)

Household Assets Value	Freq.	Percent	Cum.
0	12	10.00	10.00
20	1	0.83	10.83
60	1	0.83	11.67
100	8	6.67	18.33
140	1	0.83	19.17
200	2	1.67	20.83
220	1	0.83	21.67
260	1	0.83	22.50
300	4	3.33	25.83
340	1	0.83	26.67
400	5	4.17	30.83
450	2	1.67	32.50
500	1	0.83	33.33
550	4	3.33	36.67
590	1	0.83	37.50
650	2	1.67	39.17
670	1	0.83	40.00
700	6	5.00	45.00
770	1	0.83	45.83
800	1	0.83	46.67

850	2	1.67	48.33
890	1	0.83	49.17
900	2	1.67	50.83
950	1	0.83	51.67
1000	3	2.50	54.17
1050	1	0.83	55.00
1100	1	0.83	55.83
1200	2	1.67	57.50
1290	1	0.83	58.33
1300	1	0.83	59.17
1350	2	1.67	60.83
1450	1	0.83	61.67
1500	2	1.67	63.33
1520	1	0.83	64.17
1540	1	0.83	65.00
1550	3	2.50	67.50
1600	2	1.67	69.17
1670	1	0.83	70.00
1800	1	0.83	70.83
1900	1	0.83	71.67
1940	1	0.83	72.50
1990	1	0.83	73.33
2000	1	0.83	74.17
2010	1	0.83	75.00
2350	1	0.83	75.83
2440	1	0.83	76.67
2500	1	0.83	77.50
2550	1	0.83	78.33
2600	2	1.67	80.00
2640	1	0.83	80.83
2700	1	0.83	81.67
2950	1	0.83	82.50
2980	1	0.83	83.33
3100	1	0.83	84.17
3350	2	1.67	85.83
3700	1	0.83	86.67
4000	1	0.83	87.50
4350	1	0.83	88.33
4550	1	0.83	89.17
4660	1	0.83	90.00
4750	1	0.83	90.83
4900	1	0.83	91.67
5490	1	0.83	92.50
5560	1	0.83	93.33
6150	1	0.83	94.17
6350	1	0.83	95.00
6490	1	0.83	95.83
7090	1	0.83	96.67
7950	1	0.83	97.50
8690	1	0.83	98.33
8800	1	0.83	99.17
10650	1	0.83	100.00
Total	120	100.00	

3.11 Filter Rate of Use (Family as unit)

Rate of Use	Freq.	Percent	Cum.
0	1	1.59	1.59
.25	1	1.59	3.17
.29	2	3.17	6.35
.33	3	4.76	11.11
.4	1	1.59	12.70
.43	3	4.76	17.46
.5	6	9.52	26.98
.57	2	3.17	30.16
.75	1	1.59	31.75

1	25	39.68	71.43
1.5	3	4.76	76.19
2	13	20.63	96.83
2.5	1	1.59	98.41
3	1	1.59	100.00
Total	63	100.00	

3.12 Reason No Filter (Family as unit)

Reason No Filter	Freq.	Percent	Cum.
0	5	8.62	8.62
0&1	1	1.72	10.34
1	25	43.10	53.45
2	2	3.45	56.90
3	13	22.41	79.31
4	3	5.17	84.48
5	1	1.72	86.21
NIP	8	13.79	100.00
Total	58	100.00	

3.13 Hand Washing Habits (Family as Unit)

Wash Hands	Freq.	Percent	Cum.
0	1	0.83	0.83
1	6	5.00	5.83
1&2&3	1	0.83	6.67
1&3	1	0.83	7.50
2	7	5.83	13.33
2&3	27	22.50	35.83
2&3&4	1	0.83	36.67
2&3&5	3	2.50	39.17
2&5	7	5.83	45.00
2&6	2	1.67	46.67
3	34	28.33	75.00
3&4	1	0.83	75.83
3&5	11	9.17	85.00
3&6	2	1.67	86.67
5	14	11.67	98.33
5&6	1	0.83	99.17
6	1	0.83	100.00
Total	120	100.00	

3.14 Soap Utilization Habits (Family as unit)

Use of Soap	Freq.	Percent	Cum.
1	46	45.54	45.54
1&2	17	16.83	62.38
1&2&3	4	3.96	66.34
1&3	25	24.75	91.09
2	1	0.99	92.08
2&3	2	1.98	94.06
3	3	2.97	97.03
3&5	1	0.99	98.02

4	2	1.98	100.00
Total	101	100.00	

3.15 Visual Inspection of Soap (Family as unit)

See Soap	Freq.	Percent	Cum.
0	29	24.17	24.17
1	91	75.83	100.00
Total	120	100.00	

3.16 Sex (Individual as unit)

Sex	Freq.	Percent	Cum.
1	394	46.85	46.85
2	447	53.15	100.00
Total	841	100.00	

3.17 Age (Individual as unit)

Age	Freq.	Percent	Cum.
.1	6	0.71	0.71
.2	2	0.24	0.95
.3	2	0.24	1.19
.4	1	0.12	1.31
.5	4	0.48	1.78
.67	1	0.12	1.90
.7	1	0.12	2.02
.8	1	0.12	2.14
1	21	2.50	4.64
1.5	2	0.24	4.88
2	27	3.21	8.09
3	28	3.33	11.41
4	20	2.38	13.79
5	22	2.62	16.41
6	17	2.02	18.43
7	22	2.62	21.05
8	23	2.73	23.78
9	22	2.62	26.40
10	35	4.16	30.56
11	15	1.78	32.34
12	30	3.57	35.91
13	20	2.38	38.29
14	14	1.66	39.95
15	23	2.73	42.69
16	24	2.85	45.54
17	22	2.62	48.16
18	25	2.97	51.13
19	15	1.78	52.91
20	26	3.09	56.00
21	12	1.43	57.43
22	22	2.62	60.05
23	27	3.21	63.26
24	12	1.43	64.68
25	23	2.73	67.42
26	14	1.66	69.08
27	8	0.95	70.04
28	9	1.07	71.11
29	7	0.83	71.94
30	30	3.57	75.51

31	4	0.48	75.98
32	10	1.19	77.17
33	4	0.48	77.65
34	5	0.59	78.24
35	13	1.55	79.79
36	1	0.12	79.90
37	3	0.36	80.26
38	5	0.59	80.86
39	6	0.71	81.57
40	23	2.73	84.30
41	3	0.36	84.66
42	3	0.36	85.02
43	3	0.36	85.37
44	1	0.12	85.49
45	17	2.02	87.51
46	4	0.48	87.99
47	5	0.59	88.59
48	7	0.83	89.42
49	2	0.24	89.66
50	15	1.78	91.44
51	1	0.12	91.56
52	9	1.07	92.63
53	5	0.59	93.22
54	2	0.24	93.46
55	3	0.36	93.82
56	4	0.48	94.29
58	4	0.48	94.77
59	1	0.12	94.89
60	9	1.07	95.96
61	1	0.12	96.08
62	1	0.12	96.20
63	2	0.24	96.43
64	2	0.24	96.67
65	1	0.12	96.79
66	1	0.12	96.91
67	3	0.36	97.27
68	1	0.12	97.38
69	2	0.24	97.62
70	7	0.83	98.45
71	1	0.12	98.57
72	1	0.12	98.69
73	1	0.12	98.81
74	1	0.12	98.93
75	3	0.36	99.29
76	1	0.12	99.41
78	1	0.12	99.52
80	2	0.24	99.76
91	1	0.12	99.88
100	1	0.12	100.00
Total	841	100.00	

3.18 Education (Individual as unit)

tab ed

Education	Freq.	Percent	Cum.
0	114	15.45	15.45
1	51	6.91	22.36
10	27	3.66	26.02
11	24	3.25	29.27
12	31	4.20	33.47
13	14	1.90	35.37
15	1	0.14	35.50
17	1	0.14	35.64
2	66	8.94	44.58
21	1	0.14	44.72
23	1	0.14	44.85
3	58	7.86	52.71

39	1	0.14	52.85
4	58	7.86	60.70
41	1	0.14	60.84
5	45	6.10	66.94
6	61	8.27	75.20
7	39	5.28	80.49
8	31	4.20	84.69
9	32	4.34	89.02
NA	3	0.41	89.43
P	74	10.03	99.46
U	4	0.54	100.00
Total	738	100.00	

3.19 Education Deficit (Individual as unit)

Education Deficit	Freq.	Percent	Cum.
-1	13	1.55	1.55
0	194	23.07	24.61
1	41	4.88	29.49
2	62	7.37	36.86
3	63	7.49	44.35
4	51	6.06	50.42
5	44	5.23	55.65
6	38	4.52	60.17
7	35	4.16	64.33
8	25	2.97	67.30
9	33	3.92	71.22
10	33	3.92	75.15
11	20	2.38	77.53
12	26	3.09	80.62
13	21	2.50	83.12
14	19	2.26	85.37
15	18	2.14	87.51
16	105	12.49	100.00
Total	841	100.00	

3.20 Profession (Individual as unit)

Work	Freq.	Percent	Cum.
0	2	1.23	1.23
1	34	20.99	22.22
1&3	11	6.79	29.01
1&3&9	2	1.23	30.25
1&5	5	3.09	33.33
1&8	1	0.62	33.95
1&9	2	1.23	35.19
10	2	1.23	36.42
2	5	3.09	39.51
2&8	1	0.62	40.12
2&9	1	0.62	40.74
3	31	19.14	59.88
3&5	1	0.62	60.49
3&7	3	1.85	62.35
3&9	1	0.62	62.96
5	20	12.35	75.31
5&3	1	0.62	75.93
6	1	0.62	76.54
7	29	17.90	94.44
8	2	1.23	95.68

```
        8&9 |          1        0.62       96.30
          9 |          6        3.70      100.00
------------+-----------------------------------
      Total |        162      100.00
```

4. Bivariate Analysis

4.1 Diarrheal Incidence and Filter Ownership

4.1.1 Pearson Chi-square Test of Independence and Fisher Test

```
Some Form |
       of |      GWI Filter
  Diarheaa |         0          1 |     Total
-----------+----------------------+----------
         0 |       324        431 |       755
           |     85.26      93.49 |     89.77
-----------+----------------------+----------
         1 |        56         30 |        86
           |     14.74       6.51 |     10.23
-----------+----------------------+----------
     Total |       380        461 |       841
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =  15.3658   Pr = 0.000
           Fisher's exact =                 0.000
   1-sided Fisher's exact =                 0.000
```

4.1.2 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     380    .1473684     .018208    .3549401     .111567    .1831699
       1 |     461    .0650759    .0115006    .2469277    .0424758    .0876761
---------+--------------------------------------------------------------------
combined |     841    .1022592    .0104541    .3031692      .08174    .1227785
---------+--------------------------------------------------------------------
    diff |            .0822925    .0215359                 .040005      .12458
------------------------------------------------------------------------------
Welch's degrees of freedom:  657.123

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =   3.8212               t =   3.8212              t =   3.8212
   P < t =   0.9999         P > |t| =   0.0001          P > t =   0.0001
```

4.1.3 Two sample test of proportions

```
Two-sample test of proportion                     x: Number of obs =      380
                                                  y: Number of obs =      461

------------------------------------------------------------------------------
Variable |       Mean    Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |       .147    .0181653    8.09237   0.0000       .1113968    .1826032
       y |       .065    .0114819    5.66111   0.0000        .042496     .087504
---------+--------------------------------------------------------------------
    diff |       .082    .0214897                           .0398809    .1241191
         |  under Ho:    .0209744    3.90952   0.0001
```

```
------------------------------------------------------------------------------
              Ho: proportion(x) - proportion(y) = diff = 0

      Ha: diff < 0              Ha: diff ~= 0              Ha: diff > 0
       z =  3.910                 z =  3.910                 z =  3.910
   P < z =  1.0000          P > |z| =  0.0001          P > z =  0.0000
```

4.2 Diarrheal Incidence and Filter Ownership by Age Category

4.2.1 Age < 6 years : Pearson Chi-square Test of Independence and Fisher Test

Some Form of Diarheaa	GWI Filter 0	1	Total
0	37	58	95
	56.92	79.45	68.84
1	28	15	43
	43.08	20.55	31.16
Total	65	73	138
	100.00	100.00	100.00

```
          Pearson chi2(1) =   8.1359   Pr = 0.004
           Fisher's exact =                 0.006
   1-sided Fisher's exact =                 0.004
```

4.2.2 Age 6-15 years: Pearson Chi-square Test of Independence and Fisher Test

Some Form of Diarheaa	GWI Filter 0	1	Total
0	104	109	213
	92.86	100.00	96.38
1	8	0	8
	7.14	0.00	3.62
Total	112	109	221
	100.00	100.00	100.00

```
          Pearson chi2(1) =   8.0781   Pr = 0.004
           Fisher's exact =                 0.007
   1-sided Fisher's exact =                 0.004
```

4.2.3 Age >15 years: Pearson Chi-square Test of Independence and Fisher Test

Some Form of Diarheaa	GWI Filter 0	1	Total
0	183	264	447
	90.15	94.62	92.74
1	20	15	35
	9.85	5.38	7.26
Total	203	279	482
	100.00	100.00	100.00

```
         Pearson chi2(1) =   3.4956   Pr = 0.062
          Fisher's exact =                 0.075
  1-sided Fisher's exact =                 0.046
```

4.3 Diarrheal incidence and Quality of Housing

4.3.1 Pearson Chi-square Test of Independence

```
Some Form |
       of |                  Quality of Housing
  Diarheaa |         1        1.5          2        2.5          3 |     Total
-----------+-------------------------------------------------------+----------
         0 |       204        258        124        132         37 |       755
           |     84.30      91.17      89.21      97.06      90.24 |     89.77
-----------+-------------------------------------------------------+----------
         1 |        38         25         15          4          4 |        86
           |     15.70       8.83      10.79       2.94       9.76 |     10.23
-----------+-------------------------------------------------------+----------
     Total |       242        283        139        136         41 |       841
           |    100.00     100.00     100.00     100.00     100.00 |    100.00

          Pearson chi2(4) =  16.4236   Pr = 0.003
```

4.3.2 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     755    1.695364     .021715    .5966682    1.652735    1.737993
       1 |      86    1.482558    .0598948    .5554419    1.363471    1.601645
---------+--------------------------------------------------------------------
combined |     841    1.673603    .0205439    .5957728    1.633279    1.713926
---------+--------------------------------------------------------------------
    diff |            .2128061    .0637098                .0865374    .3390748
------------------------------------------------------------------------------
Welch's degrees of freedom:  109.153

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0                Ha: diff ~= 0                Ha: diff > 0
       t =   3.3402                 t =   3.3402                 t =   3.3402
   P < t =   0.9994          P > |t| =   0.0011          P > t =   0.0006
```

4.4 Diarrheal Incidence and Sex

4.4.1 Pearson Chi-square Test of Independence and Fisher Test

```
Some Form |
       of |          Sex
  Diarheaa |         1          2 |     Total
-----------+----------------------+----------
         0 |       357        398 |       755
           |     90.61      89.04 |     89.77
-----------+----------------------+----------
         1 |        37         49 |        86
           |      9.39      10.96 |     10.23
-----------+----------------------+----------
     Total |       394        447 |       841
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =   0.5631   Pr = 0.453
           Fisher's exact =                 0.495
```

```
   1-sided Fisher's exact =                 0.263
```

4.4.2 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |     394    .0939086    .0147144    .2920724    .0649799    .1228374
       2 |     447    .1096197    .0147933    .3127652    .0805465    .1386929
---------+--------------------------------------------------------------------
combined |     841    .1022592    .0104541    .3031692      .08174    .1227785
---------+--------------------------------------------------------------------
    diff |           -.0157111    .0208652               -.0566651     .025243
------------------------------------------------------------------------------
Welch's degrees of freedom:  838.209

                      Ho: mean(1) - mean(2) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -0.7530               t =  -0.7530              t =  -0.7530
   P < t =   0.2258          P > |t| =   0.4517          P > t =   0.7742
```

4.4.3 Two sample test of proportions

```
Two-sample test of proportion                     x: Number of obs =      394
                                                  y: Number of obs =      447

------------------------------------------------------------------------------
Variable |       Mean    Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |       .094    .0147021   6.39364   0.0000       .0651844    .1228156
       y |        .11    .0147992   7.43285   0.0000       .0809941    .1390059
---------+--------------------------------------------------------------------
    diff |      -.016    .0208607                         -.0568862    .0248862
         |  under Ho:    .0209596  -.763373   0.4452
------------------------------------------------------------------------------

               Ho: proportion(x) - proportion(y) = diff = 0

       Ha: diff < 0             Ha: diff ~= 0             Ha: diff > 0
        z = -0.763               z = -0.763               z = -0.763
    P < z =  0.2226         P > |z| =  0.4452         P > z =  0.7774
```

4.5 Diarrheal Incidence and Age

4.5.1 Pearson Chi-square Test of Independence

```
Some Form |
       of |             ageclass
  Diarheaa |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |        95        213        447 |       755
           |     68.84      96.38      92.74 |     89.77
-----------+---------------------------------+----------
         1 |        43          8         35 |        86
           |     31.16       3.62       7.26 |     10.23
-----------+---------------------------------+----------
     Total |       138        221        482 |       841
           |    100.00     100.00     100.00 |    100.00

          Pearson chi2(2) =  80.9930   Pr = 0.000
```

4.5.2 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     755    2.466225    .0257646    .7079398    2.415646    2.516804
       1 |      86    1.906977    .1028033    .9533582    1.702576    2.111377
---------+--------------------------------------------------------------------
combined |     841    2.409037    .0260473    .7553707    2.357911    2.460162
---------+--------------------------------------------------------------------
    diff |            .5592484    .1059827                .3488809     .769616
------------------------------------------------------------------------------
Welch's degrees of freedom:  96.2276

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =   5.2768               t =   5.2768               t =   5.2768
   P < t =   1.0000          P > |t| =   0.0000          P > t =   0.0000
```

4.6 Diarrheal Incidence and Sanitation Facilities

4.6.1 Pearson Chi-square Test of Independence

```
Some Form |
       of |      Sanitation (numeric)
  Diarheaa |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |       460        150        145 |       755
           |     92.93      86.71      83.82 |     89.77
-----------+---------------------------------+----------
         1 |        35         23         28 |        86
           |      7.07      13.29      16.18 |     10.23
-----------+---------------------------------+----------
     Total |       495        173        173 |       841
           |    100.00     100.00     100.00 |    100.00

          Pearson chi2(2) =  13.8346   Pr = 0.001
```

4.6.2 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     755    1.582781    .0288427    .7925182     1.52616    1.639403
       1 |      86    1.918605    .0924142    .8570142     1.73486    2.102349
---------+--------------------------------------------------------------------
combined |     841    1.617122    .0277681    .8052753    1.562619    1.671626
---------+--------------------------------------------------------------------
    diff |           -.3358232    .0968106               -.5278317   -.1438147
------------------------------------------------------------------------------
Welch's degrees of freedom:   102.66

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -3.4689               t =  -3.4689               t =  -3.4689
   P < t =   0.0004          P > |t| =   0.0008          P > t =   0.9996
```

4.7 Diarrheal Incidence and Household Size

4.7.1 Pearson Chi-square Test of Independence

Some Form of Diarheaa	Household Size 1	2	3	4	5	Total
0	3	9	20	43	83	755
	100.00	75.00	95.24	82.69	85.57	89.77
1	0	3	1	9	14	86
	0.00	25.00	4.76	17.31	14.43	10.23
Total	3	12	21	52	97	841
	100.00	100.00	100.00	100.00	100.00	100.00

Some Form of Diarheaa	Household Size 6	7	8	9	10	Total
0	61	88	110	99	102	755
	89.71	89.80	90.91	91.67	91.89	89.77
1	7	10	11	9	9	86
	10.29	10.20	9.09	8.33	8.11	10.23
Total	68	98	121	108	111	841
	100.00	100.00	100.00	100.00	100.00	100.00

Some Form of Diarheaa	Household Size 11	12	13	14	15	Total
0	38	35	25	24	15	755
	86.36	94.59	96.15	85.71	100.00	89.77
1	6	2	1	4	0	86
	13.64	5.41	3.85	14.29	0.00	10.23
Total	44	37	26	28	15	841
	100.00	100.00	100.00	100.00	100.00	100.00

Pearson chi2(14) = 14.5807 Pr = 0.407

4.7.2 Independent t-test

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	755	8.165563	.105618	2.902095	7.958223	8.372903
1	86	7.5	.3137305	2.909417	6.87622	8.12378
combined	841	8.097503	.1002799	2.908118	7.900674	8.294332
diff		.6655629	.3310317		.0092373	1.321888

Welch's degrees of freedom: 105.679

Ho: mean(0) - mean(1) = diff = 0

Ha: diff < 0 Ha: diff ~= 0 Ha: diff > 0

```
      t =   2.0106                t =   2.0106                t =   2.0106
  P < t =   0.9765          P > |t| =   0.0469            P > t =   0.0235
```

4.8 Diarrheal incidence and Educational Attainment

4.8.1 Pearson Chi-square Test of Independence

```
    Adult |
Education | Some Form of Diarheaa
  Deficit |         0          1 |     Total
```

Table Suppressed

```
    Total |       755         86 |       841 
          |    100.00     100.00 |    100.00 

        Pearson chi2(58) =  80.7794   Pr = 0.026
```

4.8.2 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     755    9.826954    .0993699    2.730412    9.631879    10.02203
       1 |      86    10.53953    .3636832    3.372659    9.816435    11.26263
---------+--------------------------------------------------------------------
combined |     841    9.899822    .0968565     2.80884    9.709712    10.08993
---------+--------------------------------------------------------------------
    diff |           -.7125813    .3770144               -1.460715    .0355525
------------------------------------------------------------------------------
Welch's degrees of freedom:  98.4106

                     Ho: mean(0) - mean(1) = diff = 0

    Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
      t =  -1.8901                t =  -1.8901                t =  -1.8901
  P < t =   0.0308          P > |t| =   0.0617            P > t =   0.9692
```

4.9 Diarrheal incidence and Source of Water

4.9.1 Pearson Chi-square Test of Independence

```
Some Form |
       of |                    Source of Water
 Diarheaa |         1        1&2      1&2&6          2          3 |     Total
-----------+-------------------------------------------------------+----------
        0 |       327         69          8        308         31 |       743 
          |     88.62      88.46     100.00      92.77      81.58 |     90.06 
-----------+-------------------------------------------------------+----------
        1 |        42          9          0         24          7 |        82 
          |     11.38      11.54       0.00       7.23      18.42 |      9.94 
-----------+-------------------------------------------------------+----------
    Total |       369         78          8        332         38 |       825 
          |    100.00     100.00     100.00     100.00     100.00 |    100.00 

          Pearson chi2(4) =   7.7424   Pr = 0.101
```

4.10 Diarrheal Incidence and Hygiene

4.10.1 Pearson Chi-square Test of Independence

```
           | Some Form of Diarheaa
       hyg |         0          1 |     Total
-----------+----------------------+----------
         1 |        16          3 |        19 
           |     84.21      15.79 |    100.00 
-----------+----------------------+----------
       1.5 |        24          7 |        31 
           |     77.42      22.58 |    100.00 
-----------+----------------------+----------
         2 |        21          4 |        25 
           |     84.00      16.00 |    100.00 
-----------+----------------------+----------
       2.5 |        38          5 |        43 
           |     88.37      11.63 |    100.00 
-----------+----------------------+----------
         3 |        49         10 |        59 
           |     83.05      16.95 |    100.00 
-----------+----------------------+----------
       3.5 |        69         13 |        82 
           |     84.15      15.85 |    100.00 
-----------+----------------------+----------
         4 |        30          3 |        33 
           |     90.91       9.09 |    100.00 
-----------+----------------------+----------
       4.5 |       123         12 |       135 
           |     91.11       8.89 |    100.00 
-----------+----------------------+----------
         5 |       118          9 |       127 
           |     92.91       7.09 |    100.00 
-----------+----------------------+----------
       5.5 |        98         10 |       108 
           |     90.74       9.26 |    100.00 
-----------+----------------------+----------
         6 |       117          7 |       124 
           |     94.35       5.65 |    100.00 
-----------+----------------------+----------
       6.5 |        43          3 |        46 
           |     93.48       6.52 |    100.00 
-----------+----------------------+----------
         7 |         9          0 |         9 
           |    100.00       0.00 |    100.00 
-----------+----------------------+----------
     Total |       755         86 |       841 
           |     89.77      10.23 |    100.00 

         Pearson chi2(12) =  18.8585   Pr = 0.092
```

4.10.2 Independent t-tests

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     755    4.535762    .0513324    1.410474     4.43499    4.636533
       1 |      86    3.918605    .1620366    1.502665    3.596433    4.240777
---------+--------------------------------------------------------------------
combined |     841    4.472652    .0493628    1.431521    4.375763     4.56954
---------+--------------------------------------------------------------------
    diff |            .6171569    .1699731                .2800638    .9542501
------------------------------------------------------------------------------
Welch's degrees of freedom:  103.217

                      Ho: mean(0) - mean(1) = diff = 0
```

```
    Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
     t =   3.6309                   t =   3.6309                  t =   3.6309
  P < t =   0.9998          P > |t| =   0.0004            P > t =   0.0002
```

4.11 Diarrheal Incidence and Assets

4.11.1 Pearson Chi-square Test of Independence

```
           | Some Form of Diarheaa
     asset |         0          1 |     Total
-----------+----------------------+----------
         1 |       167         30 |       197
           |     84.77      15.23 |    100.00
-----------+----------------------+----------
      1.25 |         5          1 |         6
           |     83.33      16.67 |    100.00
-----------+----------------------+----------
       1.5 |        25          6 |        31
           |     80.65      19.35 |    100.00
-----------+----------------------+----------
      1.75 |       173         19 |       192
           |     90.10       9.90 |    100.00
-----------+----------------------+----------
         2 |        62          5 |        67
           |     92.54       7.46 |    100.00
-----------+----------------------+----------
      2.25 |         9          0 |         9
           |    100.00       0.00 |    100.00
-----------+----------------------+----------
       2.5 |       103         15 |       118
           |     87.29      12.71 |    100.00
-----------+----------------------+----------
      2.75 |        28          0 |        28
           |    100.00       0.00 |    100.00
-----------+----------------------+----------
         3 |         8          0 |         8
           |    100.00       0.00 |    100.00
-----------+----------------------+----------
      3.25 |       110          6 |       116
           |     94.83       5.17 |    100.00
-----------+----------------------+----------
       3.5 |         8          0 |         8
           |    100.00       0.00 |    100.00
-----------+----------------------+----------
         4 |        30          4 |        34
           |     88.24      11.76 |    100.00
-----------+----------------------+----------
      4.25 |        12          0 |        12
           |    100.00       0.00 |    100.00
-----------+----------------------+----------
         5 |        15          0 |        15
           |    100.00       0.00 |    100.00
-----------+----------------------+----------
     Total |       755         86 |       841
           |     89.77      10.23 |    100.00

          Pearson chi2(13) =  22.2571   Pr = 0.052
```

4.11.2 Independent t-tests

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	755	2.18245	.0355002	.9754479	2.112759	2.252141
1	86	1.819767	.0902811	.8372321	1.640265	1.99927
combined	841	2.145363	.0333807	.968041	2.079843	2.210882
diff		.3626829	.09701		.1705008	.5548649

Welch's degrees of freedom: 113.666

Ho: mean(0) - mean(1) = diff = 0

Ha: diff < 0	Ha: diff ~= 0	Ha: diff > 0
t = 3.7386	t = 3.7386	t = 3.7386
P < t = 0.9999	P > |t| = 0.0003	P > t = 0.0001

4.12 Diarrheal Incidence and Total Coliform Content

4.12.1 Pearson Chi-square Test of Independence

Average Total Coliform	Some Form of Diarheaa 0	1	Total
Total	468	55	523
	89.48	10.52	100.00

Pearson chi2(44) = 62.6538 Pr = 0.034

4.12.2 Independent t-tests

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	468	1355.744	80.52867	1742.102	1197.5	1513.987
1	55	2114.882	254.8784	1890.229	1603.881	2625.882
combined	523	1435.576	77.47323	1771.75	1283.379	1587.774
diff		-759.1382	267.2974		-1292.867	-225.4099

Welch's degrees of freedom: 65.6578

Ho: mean(0) - mean(1) = diff = 0

Ha: diff < 0	Ha: diff ~= 0	Ha: diff > 0
t = -2.8401	t = -2.8401	t = -2.8401
P < t = 0.0030	P > |t| = 0.0060	P > t = 0.9970

4.13 Diarrheal Incidence and E.coli Content

4.13.1 Pearson Chi-square Test of Independence

Average E-Coli	Some Form of Diarheaa 0	1	Total
Total	506	57	563
	89.88	10.12	100.00

```
Pearson chi2(31) =  37.4382   Pr = 0.198
```

4.13.2 Independent t-test

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	506	192.7994	46.44719	1044.804	101.5459	284.0529
1	57	300.4035	183.2924	1383.827	-66.77521	667.5822
combined	563	203.6936	45.64072	1082.946	114.0464	293.3408
diff		-107.6041	189.0858		-485.3856	270.1774

Welch's degrees of freedom: 63.657

```
                    Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
       t =  -0.5691                 t =  -0.5691                 t =  -0.5691
   P < t =   0.2857          P > |t| =   0.5713           P > t =   0.7143
```

4.14 Diarrheal incidence and Per Capita Consumption of Filtered Water

4.14.1 Pearson Chi-square Test of Independence

pcw	Some Form of Diarheaa 0	1	Total
0	1 100.00	0 0.00	1 100.00
.7071429	1 100.00	0 0.00	1 100.00
.75	1 100.00	0 0.00	1 100.00
.855	1 100.00	0 0.00	1 100.00
.9214286	2 100.00	0 0.00	2 100.00
.9375	3 75.00	1 25.00	4 100.00
.99	1 100.00	0 0.00	1 100.00
1.0875	1 100.00	0 0.00	1 100.00
1.2375	1 100.00	0 0.00	1 100.00
1.25	2 100.00	0 0.00	2 100.00
1.29	1 100.00	0 0.00	1 100.00

1.363636	2 100.00	0 0.00	2 100.00
1.425	1 100.00	0 0.00	1 100.00
1.5	3 100.00	0 0.00	3 100.00
1.666667	2 100.00	0 0.00	2 100.00
1.875	5 83.33	1 16.67	6 100.00
2.142857	5 100.00	0 0.00	5 100.00
2.175	0 0.00	1 100.00	1 100.00
2.25	1 100.00	0 0.00	1 100.00
2.5	4 100.00	0 0.00	4 100.00
2.727273	1 100.00	0 0.00	1 100.00
3	4 100.00	0 0.00	4 100.00
3.333333	5 100.00	0 0.00	5 100.00
3.75	5 100.00	0 0.00	5 100.00
4.285714	2 100.00	0 0.00	2 100.00
5	1 100.00	0 0.00	1 100.00
5.625	1 100.00	0 0.00	1 100.00
6	3 100.00	0 0.00	3 100.00
Total	60 95.24	3 4.76	63 100.00

Pearson chi2(27) = 28.0875 Pr = 0.406

4.15 Filter Ownership and Quality of Housing

4.15.1 Pearson Chi-square Test of Independence

	Quality of Housing					
GWI Filter	1	1.5	2	2.5	3	Total
0	22 62.86	21 52.50	6 28.57	8 44.44	1 16.67	58 48.33
1	13 37.14	19 47.50	15 71.43	10 55.56	5 83.33	62 51.67

```
-----------+-------------------------------------------------------+----------
     Total |        35         40         21         18          6 |       120 
           |    100.00     100.00     100.00     100.00     100.00 |    100.00 

          Pearson chi2(4) =   9.0370   Pr = 0.060
```

4.15.2 Independent t-test

```
Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      58    1.525862     .072153    .5495007    1.381378    1.670346
       1 |      62    1.798387    .0775242    .6104263    1.643368    1.953406
---------+--------------------------------------------------------------------
combined |     120    1.666667    .0543401    .5952661    1.559068    1.774266
---------+--------------------------------------------------------------------
    diff |            -.272525    .1059059               -.4822146   -.0628355
------------------------------------------------------------------------------
Welch's degrees of freedom:  119.816

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -2.5733               t =  -2.5733               t =  -2.5733
   P < t =   0.0056          P > |t| =   0.0113          P > t =   0.9944
```

4.16 Filter Ownership and Sanitation Facilities

4.16.1 Pearson Chi-square Test of Independence

```
           |      GWI Filter
Sanitation |         0          1 |     Total
-----------+----------------------+----------
         3 |        25         43 |        68 
           |     36.76      63.24 |    100.00 
-----------+----------------------+----------
         2 |        14         12 |        26 
           |     53.85      46.15 |    100.00 
-----------+----------------------+----------
         1 |        19          7 |        26 
           |     73.08      26.92 |    100.00 
-----------+----------------------+----------
     Total |        58         62 |       120 
           |     48.33      51.67 |    100.00 

          Pearson chi2(2) =  10.3352   Pr = 0.006
```

4.16.2 Independent t-test

```
Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      58    1.896552    .1145486    .8723764    1.667172    2.125931
       1 |      62    1.419355    .0877125    .6906492    1.243963    1.594747
---------+--------------------------------------------------------------------
combined |     120        1.65    .0745199     .816325    1.502443    1.797557
---------+--------------------------------------------------------------------
    diff |            .4771969    .1442736                .1912894    .7631044
------------------------------------------------------------------------------
```

```
Welch's degrees of freedom:  110.311

                     Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
       t =   3.3076                  t =   3.3076                  t =   3.3076
   P < t =   0.9994          P > |t| =   0.0013          P > t =   0.0006
```

4.17 Filter Ownership and Asset Value

4.17.1 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      58    1.900862    .1113526    .8480365    1.677882    2.123842
       1 |      62    2.322581    .1289868    1.015643    2.064656    2.580506
---------+--------------------------------------------------------------------
combined |     120     2.11875    .0874656    .9581377    1.945559    2.291941
---------+--------------------------------------------------------------------
    diff |           -.4217186    .1704025               -.7591484   -.0842887
------------------------------------------------------------------------------
Welch's degrees of freedom:  118.456

                     Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
       t =  -2.4748                  t =  -2.4748                  t =  -2.4748
   P < t =   0.0074          P > |t| =   0.0147          P > t =   0.9926
```

4.18 Filter Ownership and Adult Education Deficit

4.18.1 Independent t-test

```
. ttest adeddef, by (gwi) unequal welch

Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      58    11.25948    .4094384     3.11819     10.4396    12.07937
       1 |      62    9.140323    .3392721    2.671431    8.461906    9.818739
---------+--------------------------------------------------------------------
combined |     120    10.16458    .2805709      3.0735    9.609025    10.72014
---------+--------------------------------------------------------------------
    diff |             2.11916    .5317381                1.065834    3.172487
------------------------------------------------------------------------------
Welch's degrees of freedom:  114.431

                     Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
       t =   3.9853                  t =   3.9853                  t =   3.9853
   P < t =   0.9999          P > |t| =   0.0001          P > t =   0.0001
```

4.19 Filter Ownership and Household Size

4.19.1 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      57    6.350877    .4389347    3.313884    5.471585    7.230169
       1 |      62     7.33871    .3355726    2.642301    6.667691    8.009729
---------+--------------------------------------------------------------------
combined |     119    6.865546    .2760239    3.011065    6.318944    7.412149
---------+--------------------------------------------------------------------
    diff |           -.9878325    .5525148               -2.082924    .1072594
------------------------------------------------------------------------------
Welch's degrees of freedom:  108.775

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -1.7879               t =  -1.7879              t =  -1.7879
   P < t =   0.0383          P > |t| =   0.0766          P > t =   0.9617
```

4.20 Filter Ownership and Sex Ratio

4.20.1 Independent t-test

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     380    1.526316    .0256477    .4999653    1.475886    1.576745
       1 |     461    1.535792    .0232528    .4992591    1.490097    1.581487
---------+--------------------------------------------------------------------
combined |     841     1.53151    .0172173    .4993031    1.497716    1.565304
---------+--------------------------------------------------------------------
    diff |            -.009476    .0346193               -.0774301    .0584781
------------------------------------------------------------------------------
Welch's degrees of freedom:  810.211

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -0.2737               t =  -0.2737              t =  -0.2737
   P < t =   0.3922          P > |t| =   0.7844          P > t =   0.6078
```

5. Multivariate Analysis

5.1 Ordinary Least Squares Regression to Model Individual Diarrhea Incidence (Linear Probability Models)

5.1.1

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  4,   836) =   15.51
       Model |  5.33313269     4  1.33328317           Prob > F      =  0.0000
    Residual |  71.8725748   836  .085971979           R-squared     =  0.0691
-------------+------------------------------           Adj R-squared =  0.0646
       Total |  77.2057075   840  .091911557           Root MSE      =  .29321

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
         gwi |   -.066872   .0210039    -3.18   0.002    -.1080986   -.0256455
         qoh |  -.0409494   .0175865    -2.33   0.020    -.0754682   -.0064305
         age |  -.0105932   .0016871    -6.28   0.000    -.0139047   -.0072818
        age2 |   .0001377   .0000246     5.59   0.000     .0000893     .000186
       _cons |   .3317124   .0361761     9.17   0.000     .2607058     .402719
------------------------------------------------------------------------------
```

5.1.2

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  5,   835) =   13.32
       Model |  5.70492903     5  1.14098581           Prob > F      =  0.0000
    Residual |  71.5007785   835  .085629675           R-squared     =  0.0739
-------------+------------------------------           Adj R-squared =  0.0683
       Total |  77.2057075   840  .091911557           Root MSE      =  .29263

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0565694   .0215373    -2.63   0.009    -.0988429   -.0142959
         qoh |  -.0261152   .0189403    -1.38   0.168    -.0632914     .011061
         san |  -.0296354   .0142223    -2.08   0.037     .0017197    .0575512
         age |  -.0106365   .0016839    -6.32   0.000    -.0139416   -.0073314
        age2 |   .0001375   .0000246     5.60   0.000     .0000893    .0001857
       _cons |   .2544323   .0517589     4.92   0.000     .1528395     .356025
------------------------------------------------------------------------------
```

5.1.3

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  6,   834) =   11.14
       Model |  5.72716067     6  .954526778           Prob > F      =  0.0000
    Residual |  71.4785468   834  .085705692           R-squared     =  0.0742
-------------+------------------------------           Adj R-squared =  0.0675
       Total |  77.2057075   840  .091911557           Root MSE      =  .29276

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |   -.054777   .0218323    -2.51   0.012    -.0976298   -.0119243
         qoh |  -.0192279   .0232792    -0.83   0.409    -.0649205    .0264648
         san |  -.0291633   .0142588    -2.05   0.041     .0011759    .0571507
         age |  -.0105902   .0016871    -6.28   0.000    -.0139016   -.0072788
        age2 |    .000137   .0000246     5.57   0.000     .0000887    .0001853
     hygiene |  -.0074816   .0146896    -0.51   0.611    -.0363146    .0213514
       _cons |   .2638879   .0550095     4.80   0.000     .1559147    .3718611
------------------------------------------------------------------------------
```

5.1.4

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  6,   834) =   11.34
       Model |  5.82493983     6  .970823305           Prob > F      =  0.0000
    Residual |  71.3807677   834   .08558845           R-squared     =  0.0754
-------------+------------------------------           Adj R-squared =  0.0688
       Total |  77.2057075   840  .091911557           Root MSE      =  .29256

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0521366   .0218551    -2.39   0.017     -.095034   -.0092392
         qoh |  -.0119077   .0224169    -0.53   0.595    -.0559079    .0320924
         san |  -.0273641   .0143477    -1.91   0.057    -.0007977     .055526
         age |  -.0105259    .001686    -6.24   0.000    -.0138353   -.0072165
        age2 |    .000136   .0000246     5.53   0.000     .0000877    .0001843
         hyg |  -.0110038   .0092927    -1.18   0.237    -.0292436     .007236
       _cons |   .2798611   .0560254     5.00   0.000     .1698938    .3898285
```

```
------------------------------------------------------------------------------
```

5.1.5

```
    Variable |       VIF       1/VIF  
-------------+----------------------
        age2 |      8.94    0.111880
         age |      8.94    0.111895
         qoh |      1.75    0.571247
         hyg |      1.74    0.575783
         san |      1.31    0.763277
         gwi |      1.16    0.860246
-------------+----------------------
    Mean VIF |      3.97
```

5.1.6

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  6,   834) =   11.37
       Model |  5.83838058     6  .973063429           Prob > F      =  0.0000
    Residual |  71.3673269   834  .085572334           R-squared     =  0.0756
-------------+------------------------------           Adj R-squared =  0.0690
       Total |  77.2057075   840  .091911557           Root MSE      =  .29253

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0525773    .021847    -2.41   0.016    -.0954589   -.0096957
       asset |  -.0092561   .0139654    -0.66   0.508    -.0366675    .0181553
         san |  -.0268358   .0143722    -1.87   0.062    -.0013741    .0550458
         age |  -.0105112   .0016849    -6.24   0.000    -.0138184    -.007204
        age2 |   .0001358   .0000246     5.53   0.000     .0000876     .000184
         hyg |  -.0101791    .009432    -1.08   0.281    -.0286924    .0083343
       _cons |    .277039   .0534568     5.18   0.000     .1721133    .3819647
```

5.1.7

```
    Variable |       VIF       1/VIF  
-------------+----------------------
         age |      8.93    0.112022
        age2 |      8.92    0.112118
       asset |      1.79    0.557393
         hyg |      1.79    0.558789
         san |      1.31    0.760532
         gwi |      1.16    0.860719
-------------+----------------------
    Mean VIF |      3.98
```

5.1.8

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  7,   833) =    9.81
       Model |   5.8791311     7  .839875872           Prob > F      =  0.0000
    Residual |  71.3265764   833  .085626142           R-squared     =  0.0761
-------------+------------------------------           Adj R-squared =  0.0684
       Total |  77.2057075   840  .091911557           Root MSE      =  .29262

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0534302   .0219203    -2.44   0.015    -.0964557   -.0104048
         qoh |  -.0164475   .0231366    -0.71   0.477    -.0618604    .0289654
         san |  -.030404   .0148509     -2.05   0.041     .0012544    .0595535
         age |  -.0104277   .0016909    -6.17   0.000    -.0137467   -.0071087
```

```
     age2 |   .0001366   .0000246     5.55   0.000     .0000883    .0001849
 fameddef |  -.0044958   .0056513    -0.80   0.427    -.0155883    .0065966
      hyg |  -.0102249   .0093461    -1.09   0.274    -.0285697    .0081198
    _cons |    .304271   .0638882     4.76   0.000     .1788702    .4296719
------------------------------------------------------------------------------

.
. test fameddef qoh hyg

 ( 1)  fameddef = 0.0
 ( 2)  qoh = 0.0
 ( 3)  hyg = 0.0

       F(  3,   833) =    1.31
            Prob > F =    0.2693
```

5.1.9

```
     Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  7,   833) =    9.83
       Model |  5.89271819     7  .841816884           Prob > F      =  0.0000
    Residual |  71.3129893   833  .085609831           R-squared     =  0.0763
-------------+------------------------------           Adj R-squared =  0.0686
       Total |  77.2057075   840  .091911557           Root MSE      =  .29259

------------------------------------------------------------------------------
   sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0540105   .0219257    -2.46   0.014    -.0970466   -.0109744
       asset |  -.0116426    .014286    -0.81   0.415    -.0396834    .0163982
         san |  -.0299399    .014894    -2.01   0.045     .0007057    .0591741
         age |  -.0104058   .0016905    -6.16   0.000    -.0137239   -.0070877
        age2 |   .0001362   .0000246     5.54   0.000      .000088    .0001845
    fameddef |   -.004462   .0056007    -0.80   0.426    -.0154552    .0065311
         hyg |  -.0095119   .0094712    -1.00   0.316    -.0281021    .0090783
       _cons |   .2991994    .060271     4.96   0.000     .1808985    .4175003
------------------------------------------------------------------------------

.
. test fameddef asset hyg

 ( 1)  fameddef = 0.0
 ( 2)  asset = 0.0
 ( 3)  hyg = 0.0

       F(  3,   833) =    1.37
            Prob > F =    0.2522
```

5.1.10

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  7,   833) =    9.94
       Model |   5.9535884     7  .850512629           Prob > F      =  0.0000
    Residual |  71.2521191   833  .085536758           R-squared     =  0.0771
-------------+------------------------------           Adj R-squared =  0.0694
       Total |  77.2057075   840  .091911557           Root MSE      =  .29247

------------------------------------------------------------------------------
   sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0578496   .0223396    -2.59   0.010    -.1016981   -.0140012
         qoh |  -.0191244   .0231698    -0.83   0.409    -.0646025    .0263537
         san |  -.0312701   .0146927    -2.13   0.034      .002431    .0601092
         age |  -.0108002   .0017003    -6.35   0.000    -.0141376   -.0074628
        age2 |   .0001403   .0000248     5.65   0.000     .0000915     .000189
     adeddef |  -.0052984   .0043203    -1.23   0.220    -.0137784    .0031816
         hyg |  -.0114254   .0092962    -1.23   0.219    -.0296722    .0068214
```

```
       _cons |   .3457452   .0776082     4.46   0.000     .1934146    .4980759
------------------------------------------------------------------------------

.
. test adeddef qoh hyg

 ( 1)  adeddef = 0.0
 ( 2)  qoh = 0.0
 ( 3)  hyg = 0.0

       F(  3,   833) =    1.60
            Prob > F =    0.1871


.
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
        age2 |      9.12    0.109701
         age |      9.09    0.109959
         qoh |      1.87    0.534400
         hyg |      1.74    0.574996
     adeddef |      1.45    0.691485
         san |      1.37    0.727411
         gwi |      1.22    0.822839
-------------+----------------------
    Mean VIF |      3.69

.
```

5.1.11

```
Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  7,   833) =    9.98
       Model |  5.97324977     7  .853321395           Prob > F      =  0.0000
    Residual |  71.2324577   833  .085513155           R-squared     =  0.0774
-------------+------------------------------           Adj R-squared =  0.0696
       Total |  77.2057075   840  .091911557           Root MSE      =  .29243

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0586726   .0223723    -2.62   0.009    -.1025852     -.01476
       asset |  -.0137606   .0144139    -0.95   0.340    -.0420525    .0145313
         san |  -.030819   .0147132    -2.09   0.037     .0019397    .0596982
         age |   -.01078  .0016979    -6.35   0.000    -.0141126   -.0074473
        age2 |   .0001399   .0000248     5.65   0.000     .0000913    .0001886
     adeddef |  -.0054175   .0043138    -1.26   0.210    -.0138847    .0030497
         hyg |  -.0105602   .0094337    -1.12   0.263    -.0290767    .0079564
       _cons |   .3415542   .0741262     4.61   0.000     .1960582    .4870502
------------------------------------------------------------------------------

.
. test adeddef asset hyg

 ( 1)  adeddef = 0.0
 ( 2)  asset = 0.0
 ( 3)  hyg = 0.0

       F(  3,   833) =    1.68
            Prob > F =    0.1697
```

5.1.12

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  6,   834) =   11.16
       Model |  5.73993611     6  .956656018           Prob > F      =  0.0000
    Residual |  71.4657714   834  .085690373           R-squared     =  0.0743
-------------+------------------------------           Adj R-squared =  0.0677
```

```
       Total |  77.2057075   840  .091911557           Root MSE      =  .29273

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0949642   .0638172    -1.49   0.137    -.2202255     .030297
         qoh |  -.0412854   .0303697    -1.36   0.174    -.1008954    .0183245
         san |  -.0283122   .0143772    -1.97   0.049     .0000924     .056532
         age |  -.0105917   .0016859    -6.28   0.000    -.0139008   -.0072826
        age2 |   .0001367   .0000246     5.55   0.000     .0000884     .000185
      gwiqoh |   .0233823   .0365827     0.64   0.523    -.0484226    .0951872
       _cons |   .2794758   .0649314     4.30   0.000     .1520276    .4069241
------------------------------------------------------------------------------

.
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
      gwiqoh |     13.37    0.074784
         gwi |      9.90    0.101011
        age2 |      8.94    0.111869
         age |      8.92    0.112045
         qoh |      3.21    0.311610
         san |      1.31    0.761051
-------------+----------------------
    Mean VIF |      7.61
```

5.1.13

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F( 10,   830) =    8.04
       Model |  6.81917178    10  .681917178           Prob > F      =  0.0000
    Residual |  70.3865357   830  .084803055           R-squared     =  0.0883
-------------+------------------------------           Adj R-squared =  0.0773
       Total |  77.2057075   840  .091911557           Root MSE      =  .29121

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0663011   .0221903    -2.99   0.003    -.1098569   -.0227453
       sanc1 |  -.0386686   .0297375    -1.30   0.194    -.0970382     .019701
       sanc2 |   .0129387   .0335268     0.39   0.700    -.0528685     .078746
       sanc3 |  (dropped)
       qohc1 |  -.0235156   .0576272    -0.41   0.683    -.1366277    .0895965
       qohc2 |  -.0828548   .0526558    -1.57   0.116    -.1862089    .0204993
       qohc3 |  -.0148733   .0533348    -0.28   0.780    -.1195603    .0898137
       qohc4 |  -.1007224   .0527428    -1.91   0.057    -.2042474    .0028026
       qohc5 |  (dropped)
         age |  -.0106471   .0016817    -6.33   0.000    -.0139481   -.0073462
        age2 |   .0001361   .0000245     5.55   0.000      .000088    .0001842
         hyg |  -.0094093   .0093643    -1.00   0.315    -.0277898    .0089711
       _cons |   .3809491   .0784456     4.86   0.000     .2269741    .5349241
------------------------------------------------------------------------------
```

5.1.14

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  6,   834) =   11.36
       Model |   5.8351296     6  .972521599           Prob > F      =  0.0000
    Residual |  71.3705779   834  .085576232           R-squared     =  0.0756
-------------+------------------------------           Adj R-squared =  0.0689
       Total |  77.2057075   840  .091911557           Root MSE      =  .29253

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0532352   .0218831    -2.43   0.015    -.0961876   -.0102827
       sanc1 |  -.0537964   .0287793    -1.87   0.062    -.1102847    .0026918
```

```
       sanc2 |  -.0105262   .0326225    -0.32   0.747    -.0745581    .0535056
       sanc3 |  (dropped)
         age |   -.010439   .0016868    -6.19   0.000    -.0137499   -.0071282
        age2 |   .0001346   .0000246     5.48   0.000     .0000864    .0001829
         hyg |  -.0138941   .0078588    -1.77   0.077    -.0293195    .0015313
       _cons |   .3507349   .0393994     8.90   0.000     .2734014    .4280685
```

5.1.15

```
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  8,   832) =    9.49
       Model |  6.45642992     8   .80705374           Prob > F      =  0.0000
    Residual |  70.7492776   832  .085035189           R-squared     =  0.0836
-------------+------------------------------           Adj R-squared =  0.0748
       Total |  77.2057075   840  .091911557           Root MSE      =  .29161

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0713815   .0216621    -3.30   0.001    -.1139003   -.0288627
       qohc1 |  -.0108618   .0571387    -0.19   0.849    -.1230148    .1012912
       qohc2 |  -.0708166   .0523799    -1.35   0.177    -.1736288    .0319956
       qohc3 |   -.014366   .0533862    -0.27   0.788    -.1191535    .0904215
       qohc4 |  -.1024689   .0527714    -1.94   0.053    -.2060497    .0011118
       qohc5 |  (dropped)
         age |   -.010684   .0016824    -6.35   0.000    -.0139864   -.0073817
        age2 |   .0001372   .0000245     5.59   0.000      .000089    .0001853
         hyg |  -.0110883   .0093011    -1.19   0.234    -.0293446     .007168
       _cons |   .3636025   .0764443     4.76   0.000     .2135561    .5136488
------------------------------------------------------------------------------
```

5.1.16

sow=="1"

```
      Source |       SS       df       MS              Number of obs =     369
-------------+------------------------------           F(  7,   361) =    4.38
       Model |  2.91100791     7  .415858272           Prob > F      =  0.0001
    Residual |  34.3085043   361  .095037408           R-squared     =  0.0782
-------------+------------------------------           Adj R-squared =  0.0603
       Total |  37.2195122   368  .101139979           Root MSE      =  .30828

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0795674   .0393941    -2.02   0.044    -.1570381   -.0020966
         qoh |  -.0384755   .0375795    -1.02   0.307    -.1123778    .0354267
         san |  -.0303973   .0243288    -1.25   0.212    -.0174467    .0782414
         age |  -.0134929   .0030222    -4.46   0.000    -.0194362   -.0075496
        age2 |   .0002044   .0000483     4.24   0.000     .0001095    .0002993
         hyg |   .0086067   .0152072     0.57   0.572    -.0212992    .0385126
     adeddef |  -.0040378   .0068128    -0.59   0.554    -.0174355    .0093598
       _cons |   .3163659   .1157353     2.73   0.007     .0887657     .543966
------------------------------------------------------------------------------
```

5.1.17

sow=="2"

```
      Source |       SS       df       MS              Number of obs =     332
-------------+------------------------------           F(  7,   324) =    4.78
       Model |  2.08343557     7  .297633653           Prob > F      =  0.0000
    Residual |  20.1816247   324  .062288965           R-squared     =  0.0936
-------------+------------------------------           Adj R-squared =  0.0740
       Total |  22.2650602   331  .067266043           Root MSE      =  .24958
```

```
------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0240818   .0288559    -0.83   0.405    -.0808504    .0326869
         qoh |  -.0618215   .0353026    -1.75   0.081    -.1312727    .0076297
         san |  -.0053138   .0211435    -0.25   0.802    -.0362821    .0469096
         age |  -.0107925   .0021903    -4.93   0.000    -.0151015   -.0064835
        age2 |   .0001225    .000029     4.22   0.000     .0000655    .0001796
         hyg |  -.0005429   .0164538    -0.03   0.974    -.0329126    .0318268
     adeddef |  -.0096226   .0062451    -1.54   0.124    -.0219087    .0026635
       _cons |   .4255946   .1163704     3.66   0.000     .1966578    .6545315
------------------------------------------------------------------------------
```

5.1.18

```
absorb(sow)
                                                      Number of obs =     825
                                                      F(  5,   815) =   12.56
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.0802
                                                      Adj R-squared =  0.0701
                                                      Root MSE      =  .28869

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0565977   .0226829    -2.50   0.013    -.1011215   -.0120738
         qoh |  -.0376296   .0223084    -1.69   0.092    -.0814182    .0061591
         age |  -.0109141   .0016751    -6.52   0.000    -.0142021   -.0076261
        age2 |     .00014   .0000244     5.73   0.000     .0000921     .000188
         hyg |  -.0032792   .0102095    -0.32   0.748    -.0233192    .0167609
       _cons |   .3375667   .0412326     8.19   0.000     .2566321    .4185012
-------------+----------------------------------------------------------------
         sow |        F(4, 815) =      1.066   0.372          (5 categories)
```

5.1.19

```
absorb(sow)
                                                      Number of obs =     825
                                                      F(  6,   814) =   10.79
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.0824
                                                      Adj R-squared =  0.0711
                                                      Root MSE      =  .28854

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0520804   .0229089    -2.27   0.023     -.097048   -.0071129
         qoh |  -.0302139   .0229433    -1.32   0.188    -.0752489    .0148211
         san |  -.0204161    .014894     1.37   0.171    -.0088191    .0496512
         age |  -.0109231   .0016742    -6.52   0.000    -.0142094   -.0076369
        age2 |   .0001396   .0000244     5.72   0.000     .0000917    .0001876
         hyg |  -.0017083   .0102682    -0.17   0.868    -.0218635    .0184469
       _cons |   .2836537   .0569666     4.98   0.000     .1718349    .3954725
-------------+----------------------------------------------------------------
         sow |        F(4, 814) =      0.978   0.419          (5 categories)

.
. test san qoh hyg

 ( 1)  san = 0.0
 ( 2)  qoh = 0.0
 ( 3)  hyg = 0.0

       F(  3,   814) =    2.49
            Prob > F =    0.0589
```

5.1.20

```
absorb (circuit)
                                                       Number of obs =     841
                                                       F(  5,   813) =    9.56
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0979
                                                       Adj R-squared =  0.0680
                                                       Root MSE      =  .29268

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0354328   .0240073    -1.48   0.140    -.0825563    .0116908
         qoh |  -.0241749   .0214736    -1.13   0.261    -.0663252    .0179754
         san |  -.0168084   .0171953     0.98   0.329    -.0169439    .0505608
         age |  -.0105303   .0017031    -6.18   0.000    -.0138732   -.0071873
        age2 |   .0001381   .0000249     5.54   0.000     .0000892     .000187
       _cons |   .2575014   .0596632     4.32   0.000     .1403893    .3746135
-------------+----------------------------------------------------------------
     circuit |       F(22, 813) =      0.985   0.481          (23 categories)
```

5.1.21

```
/*CHECK FOR PROBLEMS WITH LINEAR PROBABILITY MODEL*/
.
      Source |       SS       df       MS              Number of obs =     841
-------------+------------------------------           F(  5,   835) =   13.32
       Model |  5.70492903     5  1.14098581           Prob > F      =  0.0000
    Residual |  71.5007785   835  .085629675           R-squared     =  0.0739
-------------+------------------------------           Adj R-squared =  0.0683
       Total |  77.2057075   840  .091911557           Root MSE      =  .29263

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0565694   .0215373    -2.63   0.009    -.0988429   -.0142959
         qoh |  -.0261152   .0189403    -1.38   0.168    -.0632914     .011061
         san |  -.0296354   .0142223    -2.08   0.037     .0017197    .0575512
         age |  -.0106365   .0016839    -6.32   0.000    -.0139416   -.0073314
        age2 |   .0001375   .0000246     5.60   0.000     .0000893    .0001857
       _cons |   .2544323   .0517589     4.92   0.000     .1528395     .356025
------------------------------------------------------------------------------

.
. predict pred
(option xb assumed; fitted values)

.
. summ pred if pred<0

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
        pred |      92   -.0220973   .0142305  -.0563034  -.0004183
```

5.1.22

```
. /*CORRECT FOR HETEROSKEDASTICITY USING WLS*/
.
. reg sfodiarh gwi qoh san age age2 [aw=1/(pred*(1-pred))^0.5]
(sum of wgt is   3.0605e+03)
```

```
      Source |       SS       df       MS              Number of obs =     749
-------------+------------------------------           F(  5,   743) =    7.09
       Model |  2.80806814     5  .561613628           Prob > F      =  0.0000
    Residual |  58.8289342   743   .07917757           R-squared     =  0.0456
-------------+------------------------------           Adj R-squared =  0.0391
       Total |  61.6370023   748   .08240241           Root MSE      =  .28139

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0508014   .0217755    -2.33   0.020    -.0935503   -.0080526
         qoh |  -.0149331   .0189167    -0.79   0.430    -.0520697    .0222035
         san |  -.0273923   .0148353    -1.85   0.065    -.0017317    .0565164
         age |  -.0088298   .0019316    -4.57   0.000    -.0126217   -.0050378
        age2 |   .0001128   .0000284     3.98   0.000     .0000572    .0001685
       _cons |   .2185633   .0560182     3.90   0.000     .1085905    .3285362
------------------------------------------------------------------------------
```

5.1.23

robust

```
Regression with robust standard errors                 Number of obs =     841
                                                       F(  5,   835) =    8.33
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0739
                                                       Root MSE      =  .29263

------------------------------------------------------------------------------
             |               Robust
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0565694   .0209847    -2.70   0.007    -.0977583   -.0153805
         qoh |  -.0261152   .0200023    -1.31   0.192    -.0653759    .0131455
         san |  -.0296354   .0161776    -1.83   0.067    -.0021181     .061389
         age |  -.0106365   .0022046    -4.82   0.000    -.0149638   -.0063092
        age2 |   .0001375   .0000317     4.34   0.000     .0000753    .0001997
       _cons |   .2544323   .0619489     4.11   0.000     .1328384    .3760262
------------------------------------------------------------------------------
```

5.1.24

ageclass==1

```
      Source |       SS       df       MS              Number of obs =     138
-------------+------------------------------           F(  7,   130) =    4.45
       Model |  5.72037631     7  .817196616           Prob > F      =  0.0002
    Residual |   23.881073   130  .183700561           R-squared     =  0.1932
-------------+------------------------------           Adj R-squared =  0.1498
       Total |  29.6014493   137  .216068973           Root MSE      =   .4286

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.1607092   .0834749    -1.93   0.056    -.3258543    .0044358
         qoh |  -.1072053   .0971345    -1.10   0.272    -.2993742    .0849636
         san |  -.0484241   .0534653    -0.91   0.367    -.0573506    .1541988
         age |   -.086728   .0946124    -0.92   0.361    -.2739073    .1004513
        age2 |  -.0016275   .0170763    -0.10   0.924    -.0354109     .032156
     adeddef |  -.0102748   .0153302    -0.67   0.504    -.0406037     .020054
         hyg |  -.0042318    .033541    -0.13   0.900    -.0705886     .062125
       _cons |   .8539645   .3031229     2.82   0.006     .2542721    1.453657
------------------------------------------------------------------------------
```

5.1.25

ageclass==2

```
      Source |       SS       df       MS              Number of obs =     221
-------------+------------------------------           F(  7,   213) =    2.62
       Model |  .611726626     7  .087389518           Prob > F      =  0.0128
    Residual |  7.09868061   213  .033327139           R-squared     =  0.0793
-------------+------------------------------           Adj R-squared =  0.0491
       Total |  7.71040724   220  .035047306           Root MSE      =  .18256

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0404153   .0279252    -1.45   0.149    -.0954604    .0146298
         qoh |  -.0007068   .0273891    -0.03   0.979    -.0546952    .0532816
         san |  -.0455587   .0178672    -2.55   0.011     .0103395     .080778
         age |  -.0058282   .0364701    -0.16   0.873    -.0777168    .0660604
        age2 |  -.0000747   .0017127    -0.04   0.965    -.0034507    .0033013
     adeddef |  -.0007914   .0053265    -0.15   0.882    -.0112908     .009708
         hyg |  -.0021334   .0113984    -0.19   0.852    -.0246016    .0203348
       _cons |   .0711175   .2028714     0.35   0.726    -.3287752    .4710103
------------------------------------------------------------------------------
```

5.1.26

```
ageclass==3

      Source |       SS       df       MS              Number of obs =     482
-------------+------------------------------           F(  7,   474) =    1.25
       Model |  .588382516     7  .084054645           Prob > F      =  0.2736
    Residual |  31.8701237   474  .067236548           R-squared     =  0.0181
-------------+------------------------------           Adj R-squared =  0.0036
       Total |  32.4585062   481  .067481302           Root MSE      =   .2593

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0396885   .0261529    -1.52   0.130    -.0910785    .0117016
         qoh |   .0043946   .0269152     0.16   0.870    -.0484932    .0572824
         san |  -.0192255   .0175784    -1.09   0.275    -.0153158    .0537667
         age |   .0005638   .0033754     0.17   0.867    -.0060689    .0071964
        age2 |   4.74e-06   .0000388     0.12   0.903    -.0000716     .000081
     adeddef |  -.0066348   .0051772    -1.28   0.201     -.016808    .0035384
         hyg |  -.0114372   .0110498    -1.04   0.301    -.0331498    .0102753
       _cons |   .1473543   .1043769     1.41   0.159    -.0577442    .3524529
------------------------------------------------------------------------------
```

5.2 Probit Analysis

5.2.1

```
Iteration 0:   log likelihood = -277.54582
Iteration 1:   log likelihood = -247.36249
Iteration 2:   log likelihood = -246.43776
Iteration 3:   log likelihood = -246.43381
Iteration 4:   log likelihood = -246.43381

Probit estimates                                        Number of obs =    841
                                                        LR chi2(7)    =  62.22
                                                        Prob > chi2   = 0.0000
Log likelihood = -246.43381                             Pseudo R2     = 0.1121

------------------------------------------------------------------------------
sfodiarh |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
     gwi*|   -.056752    .021785    -2.66   0.008   .548157   -.09945 -.014054
     qoh |  -.0261903   .0220101    -1.18   0.237    1.6736  -.069329  .016949
```

```
     san |  -.0254316   .0123665    -2.05   0.040   1.61712   .001194  .049669
     age |  -.0084733   .0014137    -5.92   0.000   22.4247  -.011244 -.005702
    age2 |     .00011     .00002     5.44   0.000    822.81   .000071  .000149
     hyg |  -.0099977   .0079709    -1.25   0.211   4.47265   -.02562  .005625
 adeddef |   -.006552   .0039385    -1.65   0.099   9.89982  -.014271  .001167
---------+--------------------------------------------------------------------
  obs. P |   .1022592
 pred. P |   .0794483  (at x-bar)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| are the test of the underlying coefficient being 0
```

5.2.2

```
robust

Iteration 0:   log likelihood = -277.54582
Iteration 1:   log likelihood = -247.36249
Iteration 2:   log likelihood = -246.43776
Iteration 3:   log likelihood = -246.43381
Iteration 4:   log likelihood = -246.43381

Probit estimates                                       Number of obs =    841
                                                       Wald chi2(7)  =  52.40
                                                       Prob > chi2   = 0.0000
Log likelihood = -246.43381                            Pseudo R2     = 0.1121

------------------------------------------------------------------------------
         |               Robust
sfodiarh |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
    gwi*|   -.056752   .0199036    -2.94   0.003   .548157  -.095762 -.017742
     qoh |  -.0261903   .0226926    -1.16   0.247    1.6736  -.070667  .018286
     san |  -.0254316   .0133406    -1.89   0.059   1.61712  -.000716  .051579
     age |  -.0084733   .0015298    -5.19   0.000   22.4247  -.011472 -.005475
    age2 |     .00011   .0000217     4.88   0.000    822.81   .000067  .000153
     hyg |  -.0099977   .0078065    -1.28   0.199   4.47265  -.025298  .005303
 adeddef |   -.006552   .0041441    -1.60   0.109   9.89982  -.014674   .00157
---------+--------------------------------------------------------------------
  obs. P |   .1022592
 pred. P |   .0794483  (at x-bar)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| are the test of the underlying coefficient being 0
```

5.3 Logistic Regression

5.3.1

```
Iteration 0:   log likelihood = -277.54582
Iteration 1:   log likelihood = -249.74023
Iteration 2:   log likelihood = -245.54293
Iteration 3:   log likelihood = -245.49271
Iteration 4:   log likelihood = -245.49269

Logit estimates                                   Number of obs   =        841
                                                  LR chi2(7)      =      64.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -245.49269                       Pseudo R2       =     0.1155

------------------------------------------------------------------------------
    sfodiarh |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.7098382   .2684077    -2.64   0.008    -1.235908   -.1837688
         qoh |  -.3211465   .2979263    -1.08   0.281    -.9050714    .2627783
         san |  -.3434103   .1612332    -2.13   0.033     .0273991    .6594215
```

```
         age |  -.1139203    .018991    -6.00   0.000     -.151142   -.0766986
        age2 |   .0014449   .0002617     5.52   0.000     .0009319     .001958
         hyg |  -.1209056    .101179    -1.19   0.232    -.3192127    .0774016
     adeddef |  -.0742053   .0519816    -1.43   0.153    -.1760874    .0276768
       _cons |   .5020089   .9094715     0.55   0.581    -1.280523     2.28454
------------------------------------------------------------------------------
```

5.3.2

```
Iteration 0:   log likelihood = -277.54582
Iteration 1:   log likelihood = -249.74023
Iteration 2:   log likelihood = -245.54293
Iteration 3:   log likelihood = -245.49271
Iteration 4:   log likelihood = -245.49269

Logit estimates                                   Number of obs   =        841
                                                  Wald chi2(7)    =      50.88
                                                  Prob > chi2     =     0.0000
Log likelihood = -245.49269                       Pseudo R2       =     0.1155

------------------------------------------------------------------------------
             |               Robust
    sfodiarh |      Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.7098382    .243532    -2.91   0.004    -1.187152   -.2325243
         qoh |  -.3211465   .3128201    -1.03   0.305    -.9342627    .2919696
         san |  -.3434103   .1808985     1.90   0.058    -.0111442    .6979648
         age |  -.1139203   .0226825    -5.02   0.000    -.1583772   -.0694633
        age2 |   .0014449   .0003066     4.71   0.000     .0008441    .0020458
         hyg |  -.1209056   .0991954    -1.22   0.223     -.315325    .0735139
     adeddef |  -.0742053    .055377    -1.34   0.180    -.1827422    .0343316
       _cons |   .5020089   .8249453     0.61   0.543    -1.114854    2.118872
------------------------------------------------------------------------------
```

5.4 Ordinary Least Squares Regression on Family Per Capita Incidence of Diarrhea

5.4.1

```
      Source |       SS       df       MS              Number of obs =     120
-------------+------------------------------           F(  5,   114) =    3.59
       Model |  .364361927     5  .072872385           Prob > F      =  0.0047
    Residual |  2.31563688   114  .020312604           R-squared     =  0.1360
-------------+------------------------------           Adj R-squared =  0.0981
       Total |   2.6799988   119  .022520998           Root MSE      =  .14252

------------------------------------------------------------------------------
    famdiarh |      Coef.   Std. Err.      t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  -.0607008     .02868    -2.12   0.036    -.1175156   -.0038859
         qoh |  -.0312936   .0302103    -1.04   0.302    -.0911399    .0285527
         san |  -.0168952    .018974    -0.89   0.375    -.0206922    .0544826
         hyg |  -.0155718   .0116345    -1.34   0.183    -.0386196     .007476
     adeddef |  -.0064163   .0050141    -1.28   0.203    -.0163492    .0035166
       _cons |   .3030205   .0945378     3.21   0.002     .1157418    .4902991
------------------------------------------------------------------------------
```

5.4.2

```
      Source |       SS       df       MS              Number of obs =     120
-------------+------------------------------           F(  7,   112) =    2.96
       Model |  .418582135     7  .059797448           Prob > F      =  0.0069
    Residual |  2.26141667   112   .02019122           R-squared     =  0.1562
-------------+------------------------------           Adj R-squared =  0.1034
       Total |   2.6799988   119  .022520998           Root MSE      =   .1421

------------------------------------------------------------------------------
    famdiarh |      Coef.   Std. Err.      t     P>|t|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
         gwi |  -.0581337   .0291574    -1.99   0.049    -.1159053    -.000362
         qoh |  -.0348169   .0301972    -1.15   0.251    -.0946488     .025015
         san |  -.0149607   .0189961    -0.79   0.433    -.0226778    .0525991
         hyg |  -.0160849   .0116051    -1.39   0.168     -.039079    .0069092
     adeddef |  -.0076939   .0051111    -1.51   0.135    -.0178208    .0024331
         hhs |  -.0000999   .0177139    -0.01   0.996    -.0351977    .0349979
        hhs2 |  -.0004769   .0011624    -0.41   0.682    -.0027801    .0018263
       _cons |   .3533964   .1161596     3.04   0.003     .1232411    .5835518
------------------------------------------------------------------------------
```

5.4.3

```
gwi==0

      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  8,    33) =    3.38
       Model |  .512739916     8  .064092489           Prob > F      =  0.0061
    Residual |  .625019393    33  .018939982           R-squared     =  0.4507
-------------+------------------------------           Adj R-squared =  0.3175
       Total |  1.13775931    41  .027750227           Root MSE      =  .13762

------------------------------------------------------------------------------
    famdiarh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gwi |  (dropped)
         qoh |  -.1348252   .0575724    -2.34   0.025    -.2519571   -.0176933
         san |  -.0290926   .0320535    -0.91   0.371    -.0943061    .0361208
         hyg |  -.0142497   .0174013    -0.82   0.419    -.0496529    .0211536
     adeddef |   .0143397   .0084816     1.69   0.100    -.0029163    .0315957
         hhs |   .0282075   .0328759     0.86   0.397     -.038679    .0950941
        hhs2 |  -.0022914   .0023794    -0.96   0.343    -.0071324    .0025497
        tcav |   .0000113    .000013     0.87   0.392    -.0000152    .0000377
        ecav |   3.28e-06   .0000145     0.23   0.823    -.0000263    .0000329
       _cons |    .208584   .1908629     1.09   0.282    -.1797295    .5968974
```

5.5 OLS Regressions on Microbial Water Quality Indicators

5.5.1

```
      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  4,    37) =    1.31
       Model |  17809219.9     4  4452304.98           Prob > F      =  0.2838
    Residual |   125647862    37  3395888.16           R-squared     =  0.1241
-------------+------------------------------           Adj R-squared =  0.0295
       Total |   143457082    41  3498953.22           Root MSE      =  1842.8

------------------------------------------------------------------------------
        tcav |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         qoh |   -362.153   760.6413    -0.48   0.637    -1903.359    1179.053
         san |  -364.3213   419.8531    -0.87   0.391    -486.3818    1215.024
         hyg |  -44.87634   229.4234    -0.20   0.846    -509.7323    419.9796
     adeddef |   107.4905   108.0481     0.99   0.326    -111.4357    326.4167
       _cons |   1359.917   2259.084     0.60   0.551    -3217.422    5937.257
------------------------------------------------------------------------------
```

5.5.2

```
      Source |       SS       df       MS              Number of obs =      46
-------------+------------------------------           F(  4,    41) =    0.72
       Model |  6955971.21     4  1738992.80           Prob > F      =  0.5849
    Residual |  99372599.6    41  2423721.94           R-squared     =  0.0654
-------------+------------------------------           Adj R-squared = -0.0258
       Total |   106328571    45  2362857.13           Root MSE      =  1556.8
```

ecav	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
qoh	-292.5343	605.6521	-0.48	0.632	-1515.673	930.6048
san	-306.9452	329.5049	-0.93	0.357	-358.5035	972.3938
hyg	84.34437	180.2847	0.47	0.642	-279.7479	448.4367
adeddef	-113.1417	85.16773	-1.33	0.191	-285.1414	58.85799
_cons	1219.178	1739.101	0.70	0.487	-2293.009	4731.364

Appendix 3

Composite Indicators

```
/*CREATE COMPOSITE INCOME VARIABLE*/
. gen ass=5

. replace ass=4 if assval<=8000
(821 real changes made)

. replace ass=3 if assval<=6000
(767 real changes made)

. replace ass=2 if assval<=4000
(719 real changes made)

. replace ass=1 if assval<=2000
(603 real changes made)

. tab ass

        ass |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        603       71.70       71.70
          2 |        116       13.79       85.49
          3 |         48        5.71       91.20
          4 |         54        6.42       97.62
          5 |         20        2.38      100.00
------------+-----------------------------------
      Total |        841      100.00

.
. gen house=5

. replace house=4 if qoh==2.5
(136 real changes made)

. replace house=3 if qoh==2
(139 real changes made)

. replace house=2 if qoh==1.5
(283 real changes made)

. replace house=1 if qoh==1
(242 real changes made)

. tab house

      house |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        242       28.78       28.78
          2 |        283       33.65       62.43
          3 |        139       16.53       78.95
          4 |        136       16.17       95.12
          5 |         41        4.88      100.00
------------+-----------------------------------
      Total |        841      100.00

.
. gen asset=0

. replace asset=0.75*house + 0.25*ass
(841 real changes made)

.
. tab asset

      asset |      Freq.     Percent        Cum.
------------+-----------------------------------
```

```
          1 |        197       23.42       23.42
       1.25 |          6        0.71       24.14
        1.5 |         31        3.69       27.82
       1.75 |        192       22.83       50.65
          2 |         67        7.97       58.62
       2.25 |          9        1.07       59.69
        2.5 |        118       14.03       73.72
       2.75 |         28        3.33       77.05
          3 |          8        0.95       78.00
       3.25 |        116       13.79       91.80
        3.5 |          8        0.95       92.75
          4 |         34        4.04       96.79
       4.25 |         12        1.43       98.22
          5 |         15        1.78      100.00
------------+-----------------------------------
      Total |        841      100.00

.
. /*CREATE COMPOSITE HYGIENE VARIABLE*/
. gen hyg=0

. replace hyg=2*seesoap + hygiene
(841 real changes made)

.
. /*CREATE CATEGORY VARIABLES FOR QOH AND SAN*/
.
. tab qoh, gen(qohc)

 Quality of |
    Housing |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        242       28.78       28.78
        1.5 |        283       33.65       62.43
          2 |        139       16.53       78.95
        2.5 |        136       16.17       95.12
          3 |         41        4.88      100.00
------------+-----------------------------------
      Total |        841      100.00

.
. tab san, gen(sanc)

 Sanitation |
  (numeric) |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        495       58.86       58.86
          2 |        173       20.57       79.43
          3 |        173       20.57      100.00
------------+-----------------------------------
      Total |        841      100.00

.
. /*CREATE AGE CLASS VARIABLE*/
.
. gen ageclass=0

. replace ageclass=1 if age<=5
(138 real changes made)

. replace ageclass=2 if age>5&age<16
(221 real changes made)

. replace ageclass=3 if age>=16
(482 real changes made)
```