Distribution Agreement

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

Signature:

____________________________________ Date ________4/21/2022_________

Rachel Wallace

Associations of Water Quality and Animal Ownership with Caregiver Reported Childhood Diarrhea in Rwanda: A Baseline Analysis

By

Rachel Wallace
MPH

Hubert Department of Global Health

____
Sheela Sinharoy, PhD
Committee Chair

Associations of Water Quality and Animal Ownership with Caregiver Reported Childhood Diarrhea in Rwanda: A Baseline Analysis

By

Rachel Wallace
Bachelor of Arts in Biology
Capital University
2016

Thesis Committee Chair: Sheela Sinharoy, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Health
2022

Abstract

Associations of Water Quality and Animal Ownership with Caregiver Reported Childhood Diarrhea in Rwanda: A Baseline Analysis

By Rachel Wallace

Background/ Objective:

Despite the long history of diarrhea from identifying the disease, proposing and then identifying risk factors, continued research and interventions, diarrhea is still prevalent worldwide and has both high morbidity and mortality. The purpose of this study was to further investigate current risk factors associated with diarrhea in children < 60 months of age in Rwanda. Specifically, this study aimed to assess the association between number of E. coli Colony Forming Units per 100 mL of drinking water and caregiver reported diarrhea. This study also aimed to examine the association between animals (cows and/or chickens) observed in household compounds and caregiver reported diarrhea.

Methods:

Logistic regression models were used for quantitative data analysis, using data collected from a randomized control trial in Rwanda. Independent variables were selected based on variables commonly identified and discussed in the literature regarding childhood diarrhea. All variables from the univariable analysis that had p-values less than 0.2 were included in a multivariable regression.

Results:

There was no association between water quality of household drinking water and caregiver reported diarrhea (adjusted odds ratio: 1.00; 95% CI: 1.00, 1.00; p=0.32) for children <60 months in Eastern Province, Rwanda. Categorizing the water quality variable based on WHO water quality risk factors produced similar results, with none of the categories being statistically significantly associated with caregiver-reported diarrhea. In the univariable regression, observed animals in the household compound appeared to have a slightly protective effect on caregiver reported diarrhea (OR: 0.74; 95% CI: 0.47, 1.12; p=0.16). The adjusted odds ratio for animals observed in the household compound at baseline was 0.70 (95% CI: 0.45, 1.08; p=0.12).

Conclusion:

Current diarrhea risk factors can inform future research and programmatic endeavors to mitigate and address diarrhea in today’s world. This study found that when restricted to baseline, there was no association between drinking water quality and caregiver reported diarrhea among children in our study population. There was an inverse but not statistically significant relationship between chickens and cows observed in household compounds and caregiver reported diarrhea.

Associations of Water Quality and Animal Ownership with Caregiver Reported Childhood Diarrhea in Rwanda: A Baseline Analysis

By

Rachel Wallace
Bachelor of Arts in Biology
Capital University
2016

Thesis Committee Chair: Sheela Sinharoy, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
In Global Health
2022

Table of Contents

Table of Figures

INTRODUCTION

Diarrhea kills more children every day than AIDS, malaria, and measles together. Diarrheal diseases, for children under five, are the second leading cause of death (CDC, 2018). Specifically, diarrhea is reported as the leading cause of child mortality in Rwanda (Liu et al., 2015). Further, unsafe drinking water is ranked within the top 5 risk factors for disease in Rwanda (Forouzanfar et al., 2015). Diarrhea is a public health concern worldwide, and is especially prevalent among children in low- and middle-income countries both historically and today.

Over the past several decades, research has examined risk factors for diarrhea including water quality, and many programmatic endeavors have focused on the association between water quality and diarrhea as a means to reduce the prevalence of childhood diarrhea. Yet, despite research and programmatic endeavors and progress, diarrhea still exists as a public health concern and as a leading cause of death for children under five. Thus, more research and more programmatic endeavors are necessary to continue tackling diarrheal disease.

One strategy for mitigating childhood diarrhea is to improve drinking water quality. However, there is still much unknown about specific associations between drinking water quality and diarrhea. For instance, of the many different water quality assessments and indicators, which is the most appropriate, feasible, and accurate? Further understanding water quality measurements can lead to improved testing methods and technologies and ultimately better programming to improve water quality.

Water quality is not the only risk factor for diarrhea. Further understanding the relationship of other variables and diarrhea will also prove useful in combatting this disease. For example, animal ownership is one factor that may increase risk of diarrhea in children. Measurement of ownership of animals such as cows and chickens, and assessment of the relationship between animal ownership and childhood diarrhea, has proved difficult. Further understanding the complexity of this association with childhood diarrhea could be the link that is necessary to finally combat childhood diarrhea.

Ultimately, childhood diarrhea has plagued the world throughout many decades and centuries. Continued research, and incorporating new research findings is necessary for programming to address this disease and finally overcome it. Epidemiologists and public health professionals are still searching for data to determine diarrhea disease control strategies, determine disease risk and ultimately implement strategies to overcome childhood diarrhea (Goddard, Ban, et al., 2020).

Problem Statement

Diarrhea dates back as far as Sanskrit literature and during Hippocratic times (McMahan & DuPont, 2007). Diarrhea has been associated with crowding, war, poor sanitation, and changes in seasons throughout history (Lim & Wallace, 2004; McMahan & DuPont, 2007). Still, diarrhea is reported as the second leading cause of death in children under 5, worldwide. It accounts for 1 in 9 deaths (CDC, 2018). Despite the long history of diarrhea from identifying the disease, proposing and then identifying risk factors, continued research and interventions, it is still prevalent worldwide and has both high morbidity and mortality.

The purpose of this study was to further investigate current risk factors associated with diarrhea. Are they consistent with historical findings or have they changed as society, the environment, and the world has changed over time? Current diarrhea risk factors can inform future research and tailor programmatic endeavors to most accurately mitigate and address diarrhea in today's world.

Research Goals & Research Questions

We had two main research goals. We were interested in the association between number of E. coli Colony Forming Units (CFU) per 100 mL of drinking water and caregiver reported diarrhea. We were also interested in further understanding the association between animals (cows and/or chickens) observed in household compounds and caregiver reported diarrhea.

Hypotheses

We hypothesized there is an association between E. coli water quality levels and caregiver reported diarrhea. Specifically, we hypothesized that children in households with low levels of E. coli CFU/100 mL in their drinking water will have less caregiver reported diarrhea than children in households with higher levels of E. coli CFU/100 mL in their drinking water.

Similarly, we hypothesized there is an association between observed animals in household compounds and caregiver reported diarrhea. We hypothesized that children in households with observed animals in their household compounds will have more cases of diarrhea reported by caregivers than children who do not have animals observed in their household compounds.

Thus, our research questions were twofold. Firstly, to determine if there was an association between E. coli water quality and caregiver reported diarrhea; or if there was an association between observed animals in household compounds and caregiver reported diarrhea. Secondly, to determine the direction of that association.

Significance

The significance of this study is to identify current risk factors for diarrhea in children <60 months of age in Rwanda. These risk factors are imperative to tailoring programmatic and research endeavors in order to address the high mortality rates of diarrhea among this population, and ultimately lessen the burden of diarrheal disease in children.

LITERATURE REVIEW

Introduction to E. coli, diarrhea, and drinking water standards

Globally, diarrhea, defined by the World Health Organization (WHO) as the passage of at least three loose or liquid stools per day, remains one of the leading causes of death for children under five (World Health Oganization, 2009). In the year 2016, 5.3% of all deaths, for children under five, were diarrhea deaths (Prüss-Ustün et al., 2019). Across all ages, diarrhea was the eighth leading cause of death, and fifth leading cause among children under five in the year 2016 (Troeger et al., 2018).

There are many causes of acute diarrheal disease, including contamination of food or water with fecal microorganisms, such as Escherichia coli (E. coli). E. coli is a type of bacteria that is often used as an indicator of fecal contamination because of its prevalence (Barrag, Cuesta, & Susa, 2021). It was estimated in 2010 that 1.8 billion individuals, 28% of the global population, used unsafe water, including from drinking water sources which had fecal indicator bacteria (Onda, LoBuglio, & Bartram, 2012).

Measurement of fecal coliform bacteria is one of the main tools for detecting fecal contamination in water. Coliforms are defined as "including all of the gram-negative, nonspore forming facultatively anaerobic bacilli which ferment lactose with the production of gas within 48 hours at a temperature of 35 degrees Celsius" (Dufour, 1977). Fecal coliforms are also referred to as Thermotolerant coliforms (TTC). Another, simpler definition of TTC describes these coliforms as a class of bacteria that grow at elevated temperatures (Hodge et al., 2016). These definitions include many bacteria, such as *Escherichia, Klebsiella, Enterobacter, Citrobacter* (Dufour, 1977; Hodge et al., 2016).

E. coli is one sub-group of fecal coliforms. Specifically, within fecal coliform bacteria, E. coli is the most commonly found (Goddard, Ban, et al., 2020). E. coli is also more specific to human and animal fecal matter than other fecal coliform bacteria (Goddard, Ban, et al., 2020). Most strains of E. coli are harmless, while others can result in adverse health, including diarrhea (Prevention, 2022). Enterotoxigenic E. coli (ETEC) is a type of E. coli. It is the most frequent cause of diarrhea, especially among children (Qadri, Svennerholm, Faruque, & Sack, 2005).

Recently, studies in Latin America have further demonstrated that, for children under 5, ETEC is the most common E. Coli strain associated with diarrhea (Barrag, Cuesta, & Susa, 2021).

E. coli and fecal coliforms are the two most commonly used indicators to measure the amount of fecal contamination in water. Thus, the World Health Organization (WHO) uses E. coli and fecal coliform levels to determine drinking water quality (Goddard, Ban, et al., 2020). The WHO has published guidelines for acceptable bacteria levels found in drinking water. According to these guidelines, drinking water should not have any detectable E. coli or thermotolerant coliform bacteria in any 100 mL sample (WHO, 2017). These zero E. coli per 100 mL of water guidelines are the standard WHO guidelines for drinking water world-wide, and in all contexts, including emergency settings (World Health Oganization, 2009). SPHERE, which publishes standards for use in humanitarian settings, previously had guidelines accepting 10 TTC/100 mL as recently as 2000, until changing their guidelines to match WHO at no detectable fecal coliforms per 100 mL in 2004 (Hodge et al., 2016).

However, these guidelines are just that, "guides" to achievement. The zero E. coli per 100 mL of drinking water has not been achieved worldwide. In Rwanda, TTC contamination was found in more than 75% of houses' drinking water (Kirby et al., 2016). In fact, in Chad, Madagascar, Nigeria, Sierra Leone, and Togo greater than one third of the population were exposed >100 E. coli CFUs/100 mL in drinking water sources as recently as 2020. And, 99% of the population in Chad were consuming drinking water with detectable E. coli (Bain, Johnston, Khan, Hancioglu, & Slaymaker, 2021). Consequently, diarrhea is still prevalent. In fact, despite having established that E. coli in drinking water can lead to childhood diarrhea, both reducing E. coli levels to zero and managing diarrhea are still challenges.

Yet, despite all that is known about the association between E. coli and diarrhea, and guidelines set by the WHO defining the acceptable levels of E. Coli in drinking water, we still do not clearly understand the relationship between the amount of E. coli needed to be present to result in childhood diarrhea. This relationship between amount of something (dose) and response can be referred to as “dose response” or “exposure response”. Additionally, despite decreases seen in diarrheal disease- associated mortality and disability-adjusted life-year (DALY) rates since 1990, these reductions were not consistent across geographical regions, age categories, or sexes; this calls for special attention to specific risk factors to further tailor diarrhea solutions (Karambizi, McMahan, Blue, & Temesvari, 2021).

Risk Factors for Diarrhea: Water Quality

The world has had reductions in diarrheal disease mortality, but there is still a need to reduce the global burden of this disease. The reductions in diarrheal disease mortality is not consistent across regions, sexes, or age groups; this calls for further exploration and understanding of risk factors for diarrhea to improve diarrhea mitigation endeavors (Karambizi et al., 2021). There are many factors considered “risk factors” for diarrhea. Extensive literature has considered unsafe water and unsafe sanitation risk factors for diarrhea (Troeger et al., 2018). However, unsafe water and unsafe sanitation have inconsistent definitions and measurement tools across literature and specific studies.

Animal Ownership and Diarrhea

The F- diagram was developed by Wagner, Lanoix, et al., and shared in a 1958 World Health Organization publication (Wagner, Lanoix, & World Health, 1958). The F- diagram describes the fecal-oral pathways of disease transmission The F's that make up this diagram are: Food, Fingers, Fields, Fluids, Feces, and Flies. Animals are not explicitly included as a transmission pathway in the F-diagram. As a result, there has been in a gap in both research and programming; more attention is necessary to the role of animals as a risk for diarrhea. It has been noted that there is increased interest in the role animal feces can play along the fecal- oral pathways of diarrheal disease (Goddard, Ban, et al., 2020). However, the evidence exploring this topic is sparse, as most studies of child diarrhea have not examined animal ownership as a risk factor for diarrhea. Studies are limited in the amount of data they can collect. There are many logistical considerations studies need to consider including cost, materials needed for collecting different types of data, data analysis instruments, and training required for data collectors (Goddard, Ban, et al., 2020). These are all possible reasons why animal ownership data and/or samples from animal feces have not been collected or reported on. Those studies that have collected data on animal ownership and diarrhea have produced differing conclusions as to whether animal ownership is null, protective, or harmful towards diarrhea outcome in children (Ercumen et al., 2020; Kaur, Graham, & Eisenberg, 2017; Zambrano, Levy, Menezes, & Freeman, 2014). A systematic review and meta-analysis of animal ownership and diarrhea outcome concludes that more research is necessary to better understand the true relationship (Zambrano et al., 2014).

Dose Response

For adults, diarrhea is imminent if ETEC doses exceed 1,000,000 organisms (Barrag et al., 2021). Despite this, still, the dose response for children remains unknown (Barrag et al., 2021). Thus, even with all of the studies conducted over the years related to E. coli, diarrhea, risk factors, epidemiological characteristics, water quality testing, and creation of water quality standards, it is still unknown what the dose response between E. coli and diarrhea among children is. Systematic reviews and meta-analyses have demonstrated that water quality is a risk factor for diarrheal disease, assessed through water quality improvement and health impact studies (Clasen, Schmidt, Rabie, Roberts, & Cairncross, 2007; Fewtrell et al., 2005; Wolf et al., 2014). These studies call for re-assessing existing drinking water standards, and they have raised additional questions, such as: What is the dose response between water contamination levels and resulting diarrhea? What evidence supports the previous SPHERE standard set at 10TTC/100 mL changing to no detectable fecal coliforms per 100 mL? Is there risk of diarrhea at 10TTC/100 mL? (HODGE, 2016). A dose response analysis would answer these questions being asked by researchers. However, little data exists on how much of an ingested dose is necessary to result in infection, or diarrhea (Goddard, Ban, et al., 2020). Hodge et al.'s 2016 study proposes evidence of a dose-response relationship between diarrhea and fecal contamination of household drinking water, but this study is limited by its inability to make a causal inference (Hodge et al., 2016). Thus, it is critical that researchers continue to explore this dose-response relationship to improve public health programmatic endeavors.

Models such as quantitative microbial risk assessment (QMRA) exist to estimate the infection or disease risk from a single pathway of pathogen exposure (Goddard, Ban, et al., 2020). However, they are limited by several factors such as the quality of the data collected and used for the creation of the model (Goddard, Ban, et al., 2020). Such models make assumptions that a dose-response relationship exists between a diarrhea outcome and fecal contamination (Enger, Nelson, Clasen, Rose, & Eisenberg, 2012; Hodge et al., 2016; Howard, Pedley, & Tibatemwa, 2006). Although these models are readily used, epidemiologists and public health professionals are still searching for data to determine disease control strategies, determine disease risk and ultimately implement strategies to overcome childhood diarrhea (Goddard, Ban, et al., 2020). Determining a dose response can fill evidence gaps, standardize best practices, improve exposure assessment, and inspire a uniform data reporting system for water quality comparison across studies and geography to ultimately provide a more complex, clear, and comprehensive exposure understanding (Goddard, Ban, et al., 2020).

Diarrhea Research and Programmatic Challenges

With all of the extensive and rigorous research related to drinking water, diarrhea (specifically in children under 5), and E. coli and associated fecal pathogens, why has it been so difficult to determine risk factors and a dose response? There are numerous challenges, of which the most critical include measurement, limited data, and recall bias. Each of these challenges is discussed further below.

Measurement

Randomized control studies with humans are difficult and expensive. Animal models, which can be used in place of other human studies, do not exist for dose response E. coli and diarrhea studies. The outcomes from exposure to E. coli in animals is not consistent with the outcomes in humans (Goddard, Ban, et al., 2020). Thus, surveys and water sampling are common forms of data collection. These measurements are convenient, cost-effective, and rapid (Goddard, Ban, et al., 2020). However, they pose many disadvantages such as bias and limitations, as described further below (Goddard, Ban, et al., 2020).

Use of proxy

Methods exist to measure the concentration of E. coli in drinking water, evaluate the presence of pathogens in drinking water, and assess the volume and frequency of water consumption (Barrag et al., 2021). However, despite the existence of these methods, studies may not collect data that is accurate or comprehensive enough. For example, there is doubt that studies are designed, or data is collected in a methodical enough way, as to include all water consumed by an individual, especially water consumed outside of the household. For instance, measuring water levels only in household drinking water misses any measurement of E. coli, or the volume of consumption, related to water consumed outside of the house such as at the place of employment, in the field for farmers, at school for students, or at a friend or family member's house in or outside the same community. Thus, measuring the volume of contaminated water ingested, which ultimately must be estimated, is therefore an indirect proxy measurement (Goddard, Ban, et al., 2020). The use of a proxy is a source of error, as it may vary from actual exposure (Goddard, Ban, et al., 2020). Goddard further labels this source of measurement error as "assigning household- or community-level water quality to individuals" (Goddard, Chang, Clasen, & Sarnat, 2020).

Biological

Measurement of response to an exposure, or dose, is also challenging and complicated. There are many factors that can impact the response to an exposure of E. coli. Some of these factors are related to host susceptibility including: previous exposure, recent vaccinations, gut microbiome diversity, and pre-existing gut conditions (Goddard, Ban, et al., 2020).

Collection & Processing

There are also noted concerns regarding errors that could lead to compromised data during data collection, transport, and laboratory analysis. Many drinking water contamination and diarrhea studies utilize multiple individuals collecting water samples. This can lead to inconsistencies between sample collection methods and data collection protocol adherence. Furthermore, transporting these samples can lead to issues, especially if the temperature of the samples is not regulated, or if the length of transportation varies greatly. Finally, errors can occur due to faulty laboratory instruments as well (Goddard, Chang, Clasen, & Sarnat, 2020).

Limited data collection

Another limitation can be connected to limited data collection. Single samples of household drinking water do not comprehensively represent household water fecal contamination (Goddard, Chang, et al., 2020). Many studies are short-term and thus do not have longitudinal data collected over an extensive time period, nor associated multiple water samples.

Recall & Reporting bias

There is a recall bias associated with asking caregivers about recent childhood diarrhea. Studies have used various recall windows for reporting diarrhea, and longer recall periods have been associated with increased misrecall and misreported outcomes (Ercumen et al., 2017). To mitigate this, some researchers have recommended consistent use of a 7-day recall period to minimize recall bias (Arnold et al., 2013; Ercumen et al., 2017). Still, there are issues presented by the 7-day recall period. For example, drinking water samples collected may differ in E. coli levels than the drinking water 7- days ago that resulted in reported diarrhea (Hodge et al., 2016). This is a study design and data limitation. It does not bring questions as to whether diarrhea is associated with contaminated drinking water. In fact, there is data that re-affirms the association between E. coli contaminated drinking water resulting in increased prevalence of diarrhea (Prevalence Ratio= 1.14, 95% CI= 1.05, 1.23) (Luby et al., 2015).

METHODS

The purpose of this project is to explore different factors that may contribute to caregiver-reported diarrhea for children under five in Eastern Province, Rwanda. Logistic regression models were used for quantitative data analysis, using data collected from a randomized control trial in Rwanda.

Data Details: Population + Sample + Data Collection Methods

Data for this analysis is from a randomized control trial (RCT) led by Dr. Thomas Clasen at Emory University, with funding from the Bill and Melinda Gates Foundation. This study in Eastern Province, Rwanda evaluated the effectiveness of Lifestraw Family 2.0 water filters towards improving drinking water, specifically microbial quality. The study also assessed a pay-for-performance implementation approach ("Clasen Research Group - Emory University Rollins School of Public Health," n.d.).

Villages were eligible for inclusion in the study if they were receiving the Community-Based Environmental Health Promotion Programme (CBEHPP) in the Rwamagana district. CBEHPP is a Rwanda Ministry of Health program with goals to end open defecation, improve hygienic latrine coverage, and improve associated WASH behaviors (Sinharoy et al., 2016). The RCT took place in 60 villages, which were randomly selected from a list of all 474 eligible villages. Additionally, to be selected for the study, households within these villages had to: 1) participate in CBEHPP and 2) have either a pregnant woman living in the household or at least one child under five living in the household when the baseline survey was conducted and 3) have at least one adult (18+) who could give informed consent to participate. To select households, there was a second round of random sampling. The number of households per village was 10-72, and 25 households were randomly selected from each village using simple random sampling.

Surveys were conducted at baseline, midline, and endline. These surveys were conducted between December 2018 and September 2020. Surveys were planned to coincide with different seasons. There were two types of surveys, those specific to the individual child as well as household level surveys. 1872 total household surveys were completed and 1533 were for children <60 months at baseline. Household surveys also included drinking water microbiological samples and analysis as well as observation of household-level characteristics such as the presence of animals and the structure of the latrine. For water sampling, surveyors asked for a 100 mL sample of drinking water at each household, which was collected in a sterile container, transported to a lab, and analyzed. This data was then used to inform the water quality (*WQ, risk,*) variables. These variables are further described in Appendix Table 1.

Although there was a plethora of data collected at each survey round, this analysis focused on key variables identified at baseline. Additionally, this analysis used data collected in the RCT control group and intervention group. Because this analysis only looked at baseline data, both control and treatment groups were acceptable to include. The variables used for this analysis are listed in Appendix Table 1.

Data for this analysis was restricted to baseline only, for several reasons. Firstly, the baseline survey had some variables that were not included in follow-up surveys, such as the observation of cows or chickens in household compound. Furthermore, basic demographic variables like wealth index, determined by a variable with self-reported socioeconomic status (SES) were also only collected at baseline. Secondly, by restricting data used to baseline data, this analysis did not need to account for loss to follow up or for the effect of the intervention. Nor did this analysis need to account for the change in individual age over time, as the visits and data collection spanned over many months and several years. Additionally, only one caregiver responded to the survey at baseline, thus this analysis did not need to consider different caregivers reporting data or answering survey questions for the same individual. Data analysis was done using RStudio Cloud, a cloud-based statistical software program.

Since household data and individual data were collected via separate surveys, the first step in the analysis, once data was imported to Rstudio Cloud, was to merge the two data sets. By merging these data sets, variables from the individual surveys and household surveys were joined and attached to each individual child in the data set, including the variables for diarrhea reported by caregivers (from the individual child data set) and E. coli measurements in the household drinking water sample (from the household data set).

The next step was to examine the data set and clean variables as necessary. The first noticeable data observation was the positive (right) skewed distribution of the E. coli Colony Forming Units (CFU)/ 100 mL data (Figure 1). While the median is 72, and the mean is 215, the range of data for this variable is 0-3,000 E. coli CFU/ 100 mL. There were no Too Numerous To Count (TNTC) values. To address this skewness, and non-normal distribution, the water quality variable was adjusted. In descriptive analysis, the Williams mean, which uses the formula (Geometric Mean +1) -1 was shown. The addition of the 1 ensures that the zero values are included, as they will be given the value of 1. The subtraction of 1 after the geometric mean is taken, removes the added 1. This is one way to account for the cluster of zero values. This Williams mean value is thus the aforementioned formula applied to the water quality values taken at the household level. For this analysis, the new Williams mean variable is the Williams mean of water quality levels at baseline.

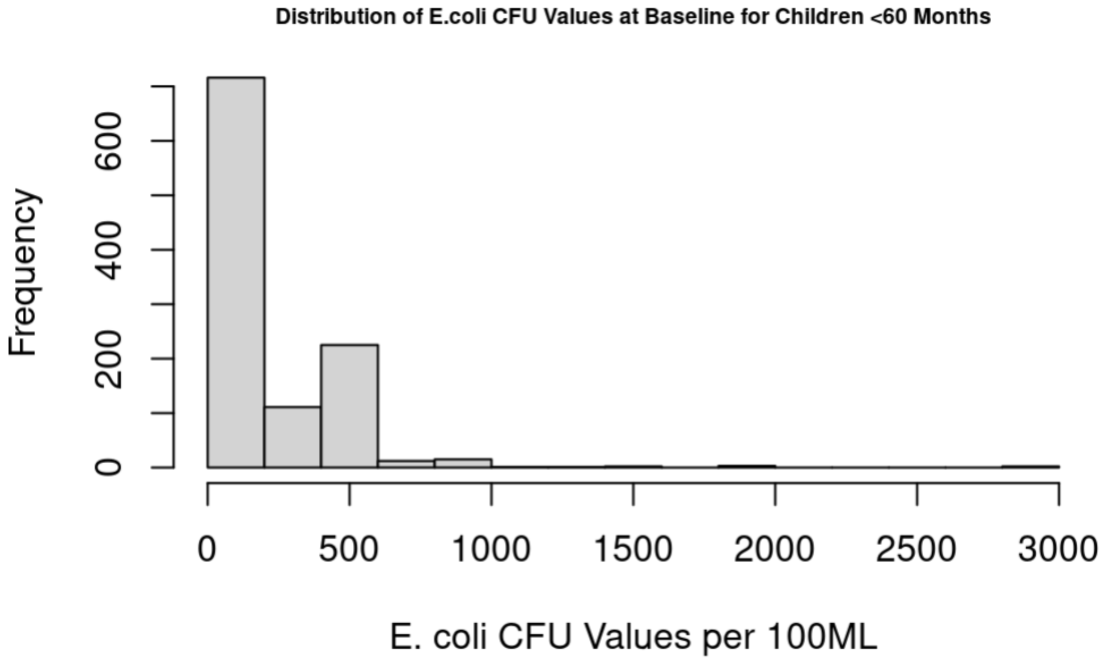


Figure 1: Histogram of Water Quality (*WQ*) Variable- E. coli CFU/100 mL at Baseline

Some variables needed to be modified so they included categories. Age in months for children, and water quality were both modified to include categories. Age in months was divided into three categories: <24 months, <60 months, 60+ months. Water quality was divided into four categories: <1, 1-10, 11-100, and 100+ CFU/100 mL, based on WHO guidelines (Organization, 1997; Sinharoy et al., 2016). The table1 package was used in R to create the descriptive characteristics tables.

Biostatistical Methods

Logistic Regression:

Logistic regressions are a type of statistical analysis, typically used to understand the relationship between a categorical dependent variable and independent variables (IBM, 2022). The dependent variable (caregiver reported diarrhea in the previous seven days) in this study was binary, making logistic regression an appropriate choice. Another reason that logistic regression was selected for this analysis is because it permits adjusting for multiple predictors (LaValley, 2008). Adjusting, in this analysis, reduces potential bias from differences in the individuals and households being compared.

The independent variables are included to look at the likelihood, or predict the relationship between these variables and the dependent variable. The logistic regression model is used to help determine the probability of what type of independent variable causes the dependent variable outcomes (IBM, 2022). For this study, we were interested in the association of caregiver reported diarrhea and many independent variables. The logistic regression outputs in RStudio Cloud are probabilities; to make the output useable and easily understandable, the outputs needed to be exponentiated. The exponentiated results can then be interpreted as an Odds Ratio (OR).

The logistic regressions using general linear model (glm) in R give odds ratio results. Odds ratios were used for this analysis because the health variable of interest, caregiver reported diarrhea, is considered a rare event, similar to Dolstad et al.'s methods ((Dolstad et al., 2021). At baseline, reported for children under 5, there were 1341 caregiver responses of no to diarrhea, 128 to yes, and 1469 total responses. Thus, the prevalence of children with caregiver reported diarrhea for the 7-day recall period was 8.7%. Since the prevalence of caregiver reported diarrhea for children under five at baseline was less than 10%, typically the threshold for a rare disease, odds ratio or relative risk can be used interchangeably (Ranganathan, Aggarwal, & Pramesh, 2015). Another study, argues that odds ratio can be used as an estimate of relative risk when the disease risk is under 20% (Cook & Sheikh, 2000). Additionally, odds ratios are commonly used in multivariate analyses, such as the multivariate analyses we have previously described for this study, when effect estimates are adjusted for factors that may differ between groups (Cook & Sheikh, 2000). Thus, we used logistic regression to calculate 95% confidence intervals, P values and odds ratios for selected variables, and we selected odds ratios because our dependent health variable is considered a rare event.

The number of observations, or data points, included in the univariable and multivariate analyses are shown in two separate "N" columns (Figures 2, 3, 4). These data points differ between variables as some values were missing or not reported in baseline surveys. The number of observations for each variable are reported in the N column with the univariable regression results. The multivariable number of observations has only one value, shown in the second N column; this value represents the “final N”, the complete data used for the complete case analysis. The number of observations deleted due to missingness in Figure 2 was 510, which results in a final N of 1023. The number of observations deleted due to missingness in Figure 3 was 512, which results in a final N of 1021. The number of observations deleted due to missingness in Figure 4 was 92, which results in a final N of 1441.

When running a logistic regression in R, one must specify a “family” term. For our analysis, family=binomial, link = “logit” was used for every analysis. This family term was used because the dependent variable was a binary variable. The link function is to constrain the predictions to fall between 0 and 1. The link function essentially transforms the linear model into a logistic regression, and ensures that the predictions are constrained to predict within the range of possible outcomes, in this case, 0 or 1 (Whalley, n.d.). The logit function is the natural log of the odds that Y equals one of the categories, for this binomial logistic regression, either 0 or 1 (Grace-Martin, 2015).

After determining the logistic regression code using glm’s family = binomial and link = logit, the models could be run. There were two main steps to this analysis. Firstly, independent variables were selected based on variables commonly identified and discussed in literature regarding childhood diarrhea. These variables included: maternal education level, household wealth (SES), age of child in months, child sex, breastfeeding history of the child, household size, and electricity available in the household (Dolstad et al., 2021; Sinharoy et al., 2016). Additionally, observed animals in the house or compound was also included as a potential independent variable. Bivariate analyses were done, in which each independent variable of interest was assessed in a logistic regression model with the dependent variable, caregiver-reported diarrhea. If the p-value in the output was less than 0.2, then these variables were selected to be in the multivariable analysis. Secondly, all variables from the univariable analysis that had been retained based on the bivariate analyses were included in a multivariable regression.

Ethics

Emory IRB (ID: MOD0004-IRB00106424) approved this study. Working with data for this study, I, Rachel Wallace, have been added as a study personnel by Emory IRB. This analysis used de-identified data, in which names of individuals and villages were replaced with unique numeric IDs. Informed consent was obtained from a parent, guardian, or caregiver in each household.

Limitations

There are limitations to this design. In the data collection methodology, there are limitations associated with self-reported surveys such as recall bias, courtesy bias, and reporting bias (Goddard, Ban, et al., 2020). Although there is inherently recall bias associated with a 7-day recall, this recall period has been recommended by some researchers in order to keep recall bias at a minimum (Arnold et al., 2013). Another limitation of this study was restricting analysis to baseline data. This was a limitation because this study collected survey data longitudinally. This particular limitation was due to analysis timelines and capacity of the research team. The data was treated as cross-sectional data rather than being examined longitudinally. Similar studies reported similar limitations (Sinharoy et al., 2016). Finally, caregiver reported diarrhea has been criticized for responder and observer bias (Rosa, Huaylinos, Gil, Lanata, & Clasen, 2014; Rosa, Kelly, & Clasen, 2016; Sinharoy et al., 2016).

RESULTS

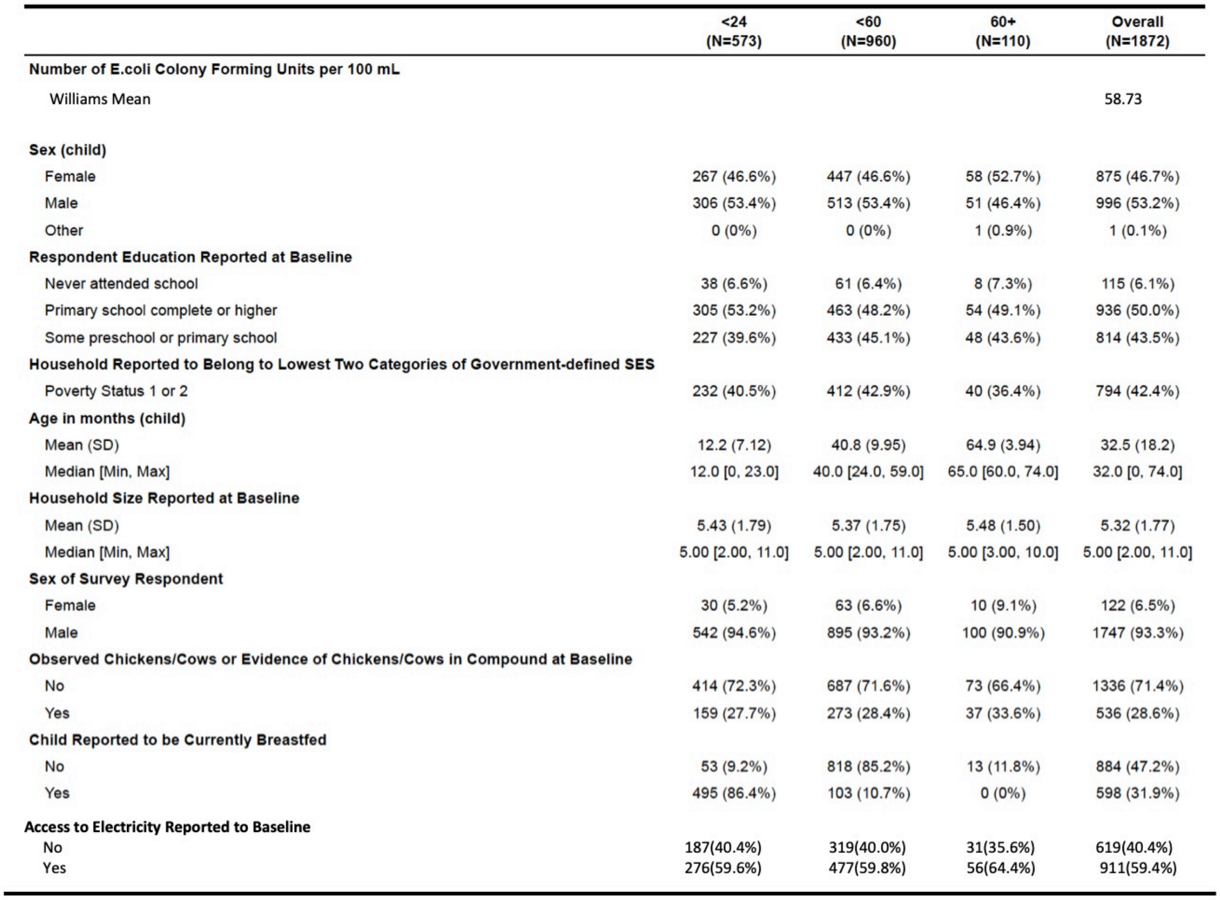
Introduction

Descriptive statistics from baseline data for socio-demographic characteristics of the study population, water quality, and diarrhea are provided in Tables 1, 2, and 3, respectively. Figures 2, 3, and 4 display findings from univariate and multivariable logistic regressions. The outcome variable was caregiver reported diarrhea. Descriptive data confirms that diarrhea was reported more often by caregivers than community health workers or than at a medical clinic. For Figure 2, the independent variable of interest was water quality. Figure 3 also had water quality as the independent variable of interest, which was divided into WHO Water Quality Risk Categories. The independent variable of interest for Figure 4 was observed cows or chickens at baseline in household compounds.

Findings

Descriptive Characteristics

As seen in Table 1, most survey respondents have attended school. Among all respondents, 936 (50%) respondents had completed primary school or higher and 814 (43.5%) respondents reported having some preschool or primary school. Additionally, 794 (42.4%) respondents reported belonging to the lowest two categories of government-defined socio-economic statuses. Furthermore, across all age categories, no chickens or cows or evidence of chickens/cows were observed at household compounds at baseline for the majority of households, 1336 (71.4%). Also, 495 (86.4%) children were reported currently breastfed for the <24 months category while the majority of <60 months aged children reported not currently breastfed 818(85.2%).

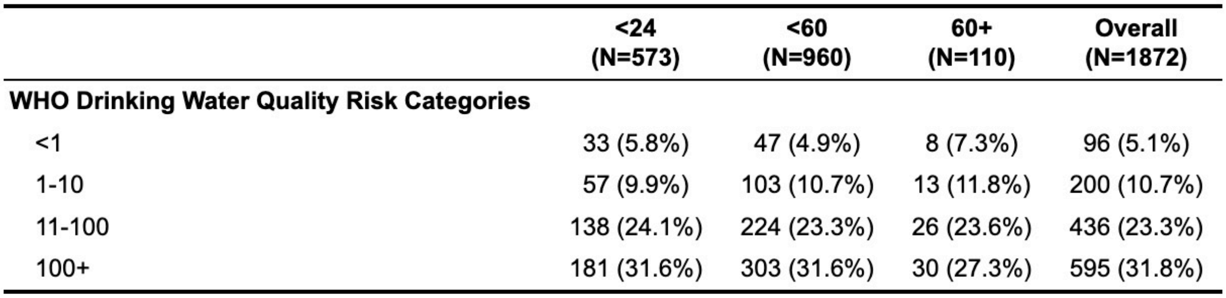


	<24 (N=573)	<60 (N=960)	60+ (N=110)	Overall (N=1872)
Number of E.coli Colony Forming Units per 100 mL				
Williams Mean				58.73
Sex (child)				
Female	267 (46.6%)	447 (46.6%)	58 (52.7%)	875 (46.7%)
Male	306 (53.4%)	513 (53.4%)	51 (46.4%)	996 (53.2%)
Other	0 (0%)	0 (0%)	1 (0.9%)	1 (0.1%)
Respondent Education Reported at Baseline				
Never attended school	38 (6.6%)	61 (6.4%)	8 (7.3%)	115 (6.1%)
Primary school complete or higher	305 (53.2%)	463 (48.2%)	54 (49.1%)	936 (50.0%)
Some preschool or primary school	227 (39.6%)	433 (45.1%)	48 (43.6%)	814 (43.5%)
Household Reported to Belong to Lowest Two Categories of Government-defined SES				
Poverty Status 1 or 2	232 (40.5%)	412 (42.9%)	40 (36.4%)	794 (42.4%)
Age in months (child)				
Mean (SD)	12.2 (7.12)	40.8 (9.95)	64.9 (3.94)	32.5 (18.2)
Median [Min, Max]	12.0 [0, 23.0]	40.0 [24.0, 59.0]	65.0 [60.0, 74.0]	32.0 [0, 74.0]
Household Size Reported at Baseline				
Mean (SD)	5.43 (1.79)	5.37 (1.75)	5.48 (1.50)	5.32 (1.77)
Median [Min, Max]	5.00 [2.00, 11.0]	5.00 [2.00, 11.0]	5.00 [3.00, 10.0]	5.00 [2.00, 11.0]
Sex of Survey Respondent				
Female	30 (5.2%)	63 (6.6%)	10 (9.1%)	122 (6.5%)
Male	542 (94.6%)	895 (93.2%)	100 (90.9%)	1747 (93.3%)
Observed Chickens/Cows or Evidence of Chickens/Cows in Compound at Baseline				
No	414 (72.3%)	687 (71.6%)	73 (66.4%)	1336 (71.4%)
Yes	159 (27.7%)	273 (28.4%)	37 (33.6%)	536 (28.6%)
Child Reported to be Currently Breastfed				
No	53 (9.2%)	818 (85.2%)	13 (11.8%)	884 (47.2%)
Yes	495 (86.4%)	103 (10.7%)	0 (0%)	598 (31.9%)
Access to Electricity Reported to Baseline				
No	187(40.4%)	319(40.0%)	31(35.6%)	619(40.4%)
Yes	276(59.6%)	477(59.8%)	56(64.4%)	911(59.4%)

Table 1: Descriptive Child, Respondent, & Household Characteristics

Table 2 shows descriptive characteristics for household drinking water samples collected at baseline. The WHO Drinking Water Quality Risk Categories are <1, 1-100, 11-100, 100+ CFU/100 mL. Most drinking water samples fall into the water quality category 100+, overall

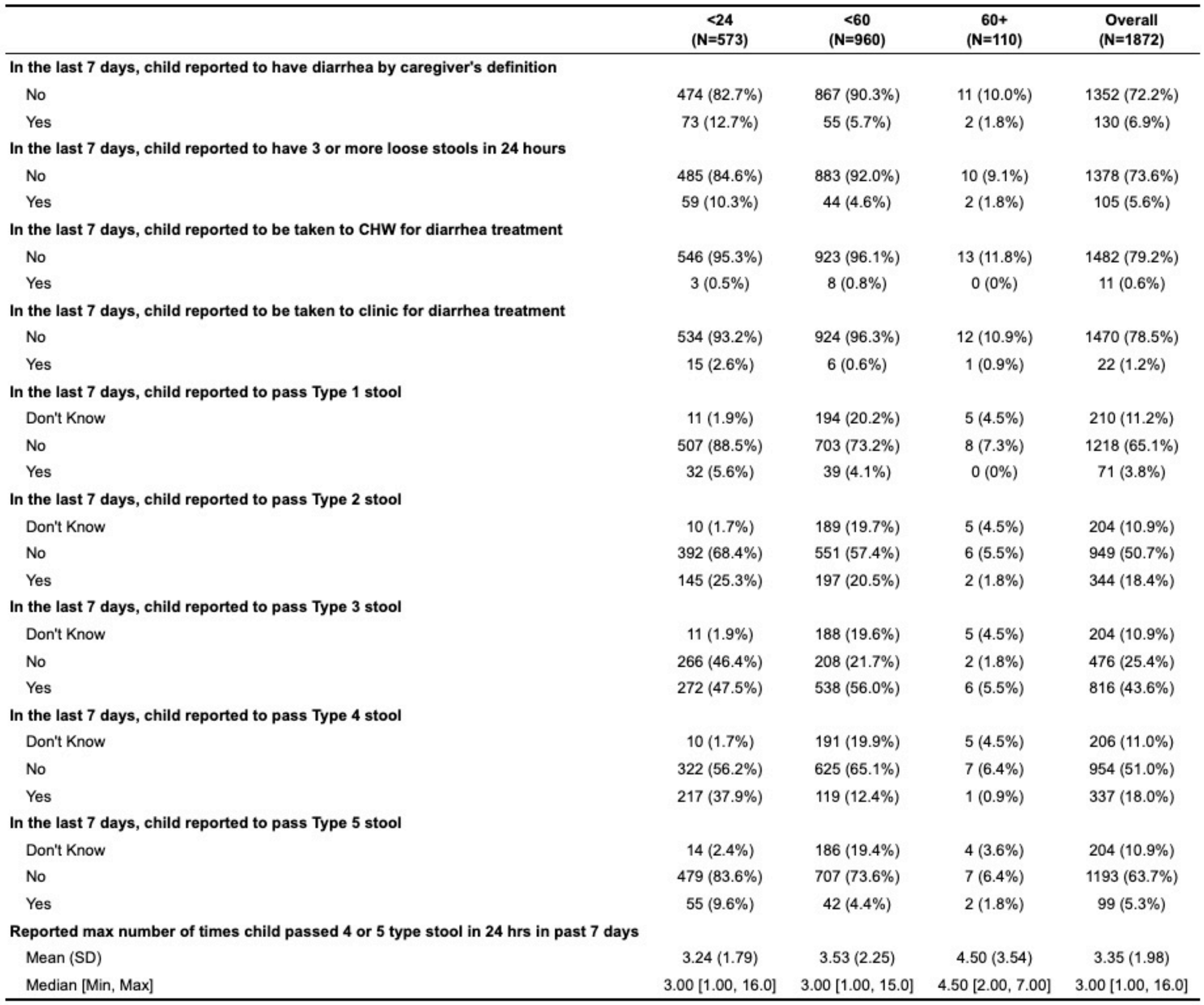
595(31.8%). Across both <24 months 181(31.6%) and <60 months 303(31.6%) and 60+ months 30(27.3%), the 100+ CFU/100 mL category has the most household samples. Notably, there were many households without household drinking water quality data for baseline. In fact, overall, there were 545 (29.1%) households missing this data. The fewest household drinking water samples fall into the <1 CFU/100 mL category. Overall, only 96 (5.1%) households had drinking water that fell into this risk category.



	<24 (N=573)	<60 (N=960)	60+ (N=110)	Overall (N=1872)
WHO Drinking Water Quality Risk Categories				
<1	33 (5.8%)	47 (4.9%)	8 (7.3%)	96 (5.1%)
1-10	57 (9.9%)	103 (10.7%)	13 (11.8%)	200 (10.7%)
11-100	138 (24.1%)	224 (23.3%)	26 (23.6%)	436 (23.3%)
100+	181 (31.6%)	303 (31.6%)	30 (27.3%)	595 (31.8%)

Table 2: Water Quality Categories Descriptive Characteristics

Diarrhea characteristics are shown in Table 3. The percentage of children reported to have diarrhea by the caregiver's definition was 6.9%. This percentage was higher than the percentages of children reported to have diarrhea based on other methods of assessment. Type 3 stool was the most commonly reported type of stool, reported for 816 (43.6%) overall, and across both <24 months 272 (47.5%) and <60 months 538 (56.0%) age groups. Type 1 and Type 2 stool reported, in the defined 7-day recall period, had very high "Don't Know" responses 210 (11.2%) and 204 (10.9%) respectively. Children <24 months had higher prevalence of diarrhea than children <60 months based on almost all assessment methods. Very few respondents reported seeking treatment for a child for diarrhea: among all respondents overall, only 11 (0.6%) and 22 (1.2%) children had been taken to a CHW or clinic for diarrhea treatment, respectively.

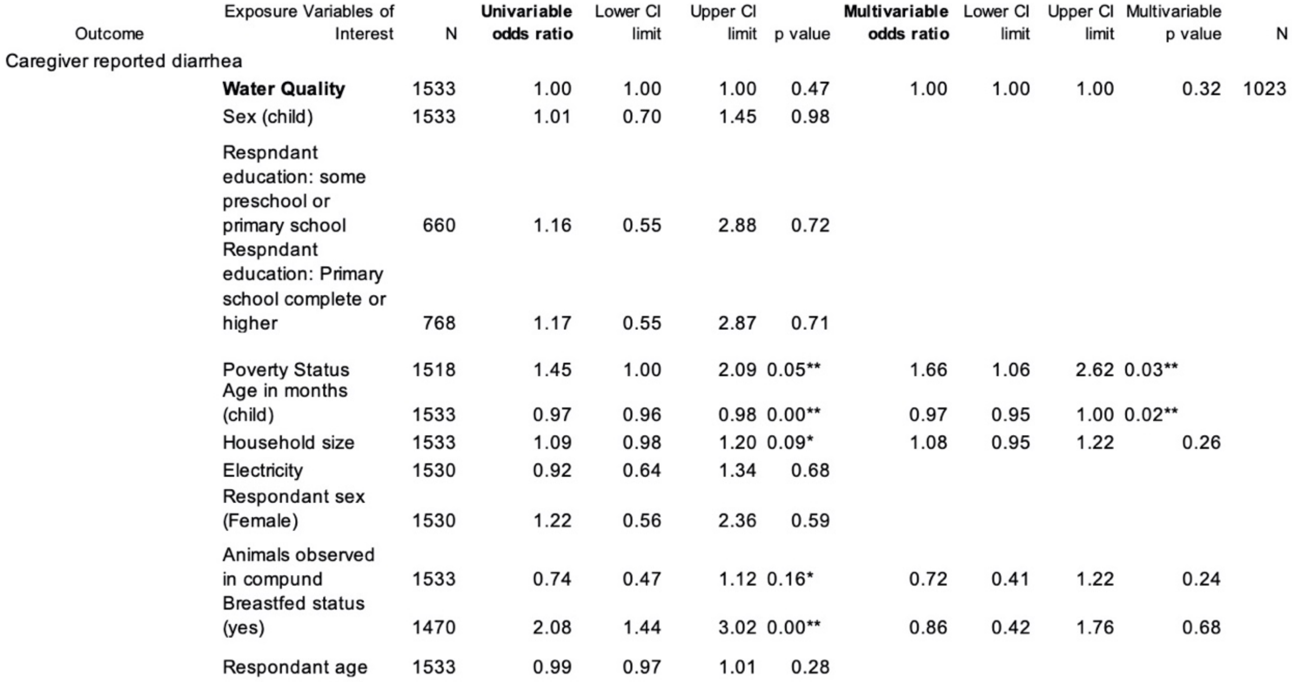


	<24 (N=573)	<60 (N=960)	60+ (N=110)	Overall (N=1872)
In the last 7 days, child reported to have diarrhea by caregiver's definition				
No	474 (82.7%)	867 (90.3%)	11 (10.0%)	1352 (72.2%)
Yes	73 (12.7%)	55 (5.7%)	2 (1.8%)	130 (6.9%)
In the last 7 days, child reported to have 3 or more loose stools in 24 hours				
No	485 (84.6%)	883 (92.0%)	10 (9.1%)	1378 (73.6%)
Yes	59 (10.3%)	44 (4.6%)	2 (1.8%)	105 (5.6%)
In the last 7 days, child reported to be taken to CHW for diarrhea treatment				
No	546 (95.3%)	923 (96.1%)	13 (11.8%)	1482 (79.2%)
Yes	3 (0.5%)	8 (0.8%)	0 (0%)	11 (0.6%)
In the last 7 days, child reported to be taken to clinic for diarrhea treatment				
No	534 (93.2%)	924 (96.3%)	12 (10.9%)	1470 (78.5%)
Yes	15 (2.6%)	6 (0.6%)	1 (0.9%)	22 (1.2%)
In the last 7 days, child reported to pass Type 1 stool				
Don't Know	11 (1.9%)	194 (20.2%)	5 (4.5%)	210 (11.2%)
No	507 (88.5%)	703 (73.2%)	8 (7.3%)	1218 (65.1%)
Yes	32 (5.6%)	39 (4.1%)	0 (0%)	71 (3.8%)
In the last 7 days, child reported to pass Type 2 stool				
Don't Know	10 (1.7%)	189 (19.7%)	5 (4.5%)	204 (10.9%)
No	392 (68.4%)	551 (57.4%)	6 (5.5%)	949 (50.7%)
Yes	145 (25.3%)	197 (20.5%)	2 (1.8%)	344 (18.4%)
In the last 7 days, child reported to pass Type 3 stool				
Don't Know	11 (1.9%)	188 (19.6%)	5 (4.5%)	204 (10.9%)
No	266 (46.4%)	208 (21.7%)	2 (1.8%)	476 (25.4%)
Yes	272 (47.5%)	538 (56.0%)	6 (5.5%)	816 (43.6%)
In the last 7 days, child reported to pass Type 4 stool				
Don't Know	10 (1.7%)	191 (19.9%)	5 (4.5%)	206 (11.0%)
No	322 (56.2%)	625 (65.1%)	7 (6.4%)	954 (51.0%)
Yes	217 (37.9%)	119 (12.4%)	1 (0.9%)	337 (18.0%)
In the last 7 days, child reported to pass Type 5 stool				
Don't Know	14 (2.4%)	186 (19.4%)	4 (3.6%)	204 (10.9%)
No	479 (83.6%)	707 (73.6%)	7 (6.4%)	1193 (63.7%)
Yes	55 (9.6%)	42 (4.4%)	2 (1.8%)	99 (5.3%)
Reported max number of times child passed 4 or 5 type stool in 24 hrs in past 7 days				
Mean (SD)	3.24 (1.79)	3.53 (2.25)	4.50 (3.54)	3.35 (1.98)
Median [Min, Max]	3.00 [1.00, 16.0]	3.00 [1.00, 15.0]	4.50 [2.00, 7.00]	3.00 [1.00, 16.0]

Table 3: Diarrhea Characteristics

Logistic Regressions

Figures 2, 3, and 4 show the crude and adjusted log-binomial regression model results. In these models, the outcome was caregiver-reported diarrhea for children under five. In Figure 2, the primary independent variable of interest was water quality. Figure 3 further breaks down water quality into WHO risk categories. These risk categories are the primary independent variable of interest for Figure 3. In Figure 4, the primary independent variable of interest was observed animals in the family compound.



Outcome	Exposure Variables of Interest	N	**Univariable odds ratio**	Lower CI limit	Upper CI limit	p value	**Multivariable odds ratio**	Lower CI limit	Upper CI limit	Multivariable p value	N
Caregiver reported diarrhea											
	Water Quality	1533	1.00	1.00	1.00	0.47	1.00	1.00	1.00	0.32	1023
	Sex (child)	1533	1.01	0.70	1.45	0.98					
	Respndant education: some preschool or primary school	660	1.16	0.55	2.88	0.72					
	Respndant education: Primary school complete or higher	768	1.17	0.55	2.87	0.71					
	Poverty Status	1518	1.45	1.00	2.09	0.05**	1.66	1.06	2.62	0.03**	
	Age in months (child)	1533	0.97	0.96	0.98	0.00**	0.97	0.95	1.00	0.02**	
	Household size	1533	1.09	0.98	1.20	0.09*	1.08	0.95	1.22	0.26	
	Electricity	1530	0.92	0.64	1.34	0.68					
	Respondant sex (Female)	1530	1.22	0.56	2.36	0.59					
	Animals observed in compund	1533	0.74	0.47	1.12	0.16*	0.72	0.41	1.22	0.24	
	Breastfed status (yes)	1470	2.08	1.44	3.02	0.00**	0.86	0.42	1.76	0.68	
	Respondant age	1533	0.99	0.97	1.01	0.28					

* = p value less than 0.2   ** = p value less than 0.05

Figure 2: Logistic Regression with Outcome Caregiver Reported Diarrhea and Independent Variable of Interest Water Quality

Univariable regressions

Among the univariable regressions, with diarrhea as the dependent variable (or outcome), five variables had p values of less than 0.05 or less than 0.2, as shown in Figure 2. These variables were included as covariates in all of the multivariable regressions. Poverty status (i.e., households reported to belong to the lowest two categories of government-defined SES) was positively associated with caregiver-reported diarrhea. The odds of caregiver reported diarrhea are 1.45 (95% CI= 1.00, 2.09) times higher among those households that report belonging to the lowest two SES categories as opposed to those households who do not belong to these SES categories. For every 1-person increase in the household size, the odds of caregiver reported diarrhea are 1.09 (95% CI= 0.98, 1.20) higher. The odds of caregiver reported diarrhea are 2.08

(95% CI= 1.44, 3.02) times higher among those who were breastfed compared to those who were not breastfed.

In contrast, some variables appeared to have inverse association with caregiver reported diarrhea. For every 1-month increase in the child's age in months, the odds of caregiver reported diarrhea decrease by 3% (OR=0.97, 95% CI= 0.96, 0.98). The odds of caregiver reported diarrhea are 26% (OR= 0.74, 95% CI= 0.47, 1.12) lower among those who have animals observed in the household compound compared to those who do not.

Multivariable regressions

Water Quality

As shown in Figure 2, when adjusted for SES status, age of the child, household size, animals observed in the household compound, and breastfed status of the child, the odds ratio of drinking water quality remained 1.00 (95% CI= 1.00,1.00) and the p value (p=0.32), although reduced from the univariable p value, remained non-significant. This odds ratio at 1.00 suggests that there is truly no association between water quality and caregiver reported diarrhea in this study sample.

Outcome	Exposure Variables of Interest	N	**Univariable odds ratio**	Lower CI limit	Upper CI limit	p value	**Multivariable odds ratio**	Lower CI limit	Upper CI limit	Multivariable p value	N
Caregiver reported diarrhea											
	Water Quality Categories	1086									1021
	<1	80									
	1-10	160	0.67	0.25	1.93	0.44	0.70	0.25	2.03	0.49	
	11-100	362	1.05	0.47	2.66	0.92	1.04	0.46	2.67	0.94	
	100+	484	0.79	0.36	2.01	0.60	0.75	0.33	1.92	0.51	
	Poverty Status	1518	1.45	1.00	2.09	0.05**	1.65	1.05	2.61	0.03**	
	Age in months (child)	1533	0.97	0.96	0.98	0.00**	0.97	0.95	1.00	0.03**	
	Household size	1533	1.09	0.98	1.20	0.09*	1.08	0.95	1.22	0.26	
	Animals observed in compund	1533	0.74	0.47	1.12	0.16*	0.71	0.40	1.19	0.21	
	Breastfed status (yes)	1470	2.08	1.44	3.02	0.00**	0.88	0.43	1.82	0.74	

* = p value less than 0.2 ** = p value less than 0.05

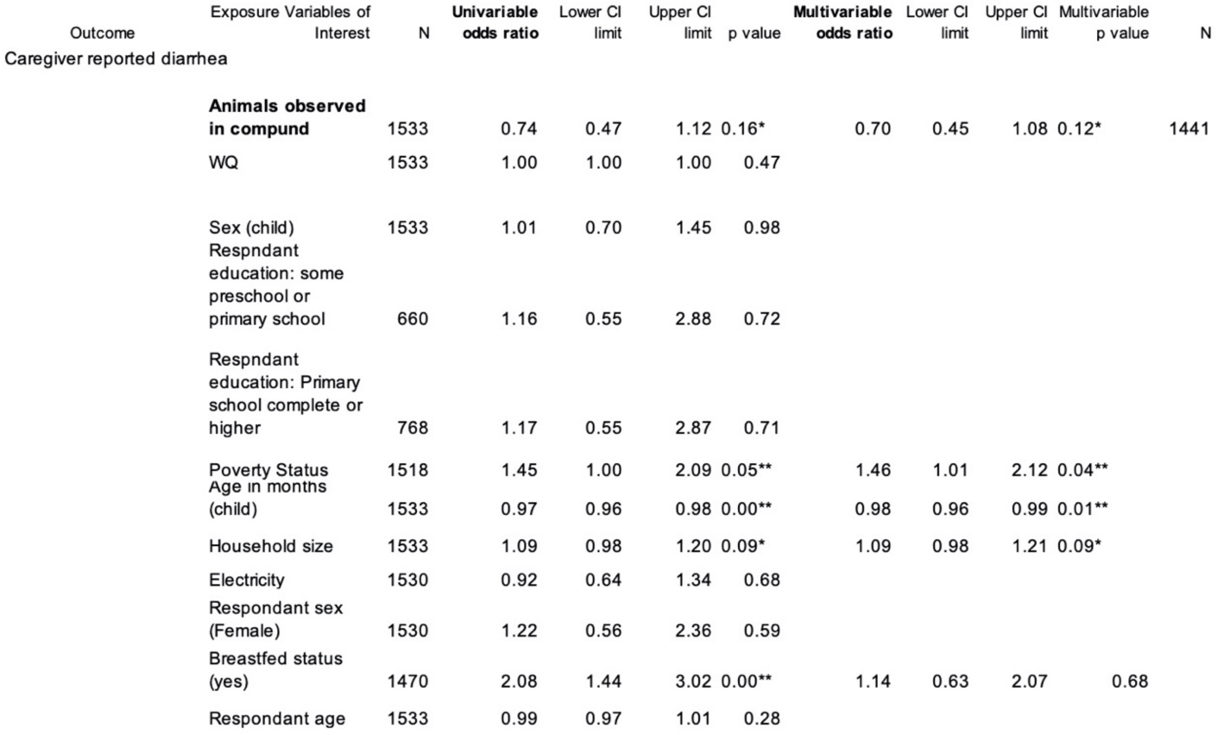
Figure 3: Logistic Regression with Outcome Caregiver Reported Diarrhea and Independent Variable of Interest Water Quality, Categorized by WHO Water Quality Risk Groups

Water Quality Categories

As shown in Figure 3, when compared to the water quality category with less than 1 E. coli Colony Forming Units/100 mL, all of the other water quality categories had non-significant p values and 95% confidence intervals that included 1.00 both in adjusted and unadjusted models. As with the continuous water quality variable, this result indicated no association between water quality and caregiver reported diarrhea in this study sample.

Observations of Animals within Household Compounds

As shown in Figure 4, when adjusted for SES status, age of the child, household size, and breastfed status of the child, the odds of caregiver reported diarrhea is 30% (OR=0.70, 95% CI= 0.45,1.08) lower among those who have animals observed in the household compound compared to those who do not. However, the p value (p=0.12) was not statistically significant.



Outcome	Exposure Variables of Interest	N	**Univariable odds ratio**	Lower CI limit	Upper CI limit	p value	**Multivariable odds ratio**	Lower CI limit	Upper CI limit	Multivariable p value	N
Caregiver reported diarrhea											
	Animals observed in compund	1533	0.74	0.47	1.12	0.16*	0.70	0.45	1.08	0.12*	1441
	WQ	1533	1.00	1.00	1.00	0.47					
	Sex (child)	1533	1.01	0.70	1.45	0.98					
	Respndant education: some preschool or primary school	660	1.16	0.55	2.88	0.72					
	Respndant education: Primary school complete or higher	768	1.17	0.55	2.87	0.71					
	Poverty Status	1518	1.45	1.00	2.09	0.05**	1.46	1.01	2.12	0.04**	
	Age in months (child)	1533	0.97	0.96	0.98	0.00**	0.98	0.96	0.99	0.01**	
	Household size	1533	1.09	0.98	1.20	0.09*	1.09	0.98	1.21	0.09*	
	Electricity	1530	0.92	0.64	1.34	0.68					
	Respondant sex (Female)	1530	1.22	0.56	2.36	0.59					
	Breastfed status (yes)	1470	2.08	1.44	3.02	0.00**	1.14	0.63	2.07	0.68	
	Respondant age	1533	0.99	0.97	1.01	0.28					

* = p value less than 0.2 ** = p value less than 0.05

Figure 4: Logistic Regression with Outcome Caregiver Reported Diarrhea and Independent Variable of Interest Animals Observed in Household Compound

Summary

Water quality's adjusted odds ratio of 1.00 (95% CI= 1.00, 1.00) and p value of 0.32 demonstrated that there is no association between water quality of household drinking water and caregiver reported diarrhea for children under 5 (<60months) in this population in Rwanda. Categorizing the water quality variable based on WHO Water Quality Risk Factors produced similar results, with none of the categories being statistically significantly associated with caregiver-reported diarrhea. Furthermore, in the univariable regression, observed animals in the household compound appeared to have a slightly protective effect on caregiver reported diarrhea

0.74 (95% CI = 0.47, 1.12) and p value of 0.16. The adjusted odds ratio for animals observed in the household compound at baseline was 0.70 (95% CI = 0.45, 1.08) and a p value of 0.12.

DISCUSSION

Introduction

Ultimately, the logistic regression results showed that, at baseline, there was no association between water quality and caregiver reported diarrhea as a whole and when dividing into WHO risk categories. Animals observed in the household compound had an inverse association with caregiver reported diarrhea, although this result was not statistically significant.

Discussion

This study did not find drinking water quality, as measured by E. coli CFU, to be a risk factor for childhood diarrhea. This is consistent with findings from another baseline paper from Rwanda, which found no association between CFUs of thermotolerant (fecal) coliforms and caregiver-reported diarrhea (Sinharoy et al., 2016). E. coli CFU was not found to be a risk factor for childhood diarrhea in several other studies as well. A systematic review looking at E. coli or fecal coliforms in drinking water as risk factors for diarrhea did not find conclusive evidence of an association between E. coli or fecal coliforms and diarrhea (Risk Ratio=1.26, 95% CI= 1.37, 1.74) (Gruber, Ercumen, & Colford, 2014). Another study researched E. coli contamination levels among diarrhea transmission pathways: source water, stored water, pond water, child hands, soil, food, flies, and only determined E. coli levels on children's hands to be significantly associated with child diarrhea (Pickering et al., 2018). Though, there does appear to be variability in evidence and findings for E. coli and diarrhea's association. Luby et al. did find an association between E. coli and diarrhea though described the E. coli drinking water contamination and child diarrhea association's strength varied by child age; with children aged 6 to <12 months having strongest association. And, findings described high variability in E. coli levels in households between study visits (Luby et al., 2015). Ercumen et al. also concluded finding evidence which supported a strong relationship between E. coli and diarrhea, this study focused on eliminating bias in their study, including using a 2-day recall period (Ercumen et al., 2017).

Using different measures, other studies did find household drinking water to be a determinant for childhood diarrhea. Findings from a community-based, case control study in Ethiopia determined unprotected drinking water (Adjusted OR=1.83, 95% CI= 1.12, 2.98) to be a significant factor in childhood diarrhea (Asfaha et al., 2018). However, that study did not use E. coli as the indicator of drinking water status, instead looking at protected versus unprotected drinking water source. Another study that determined household drinking water to be a determinant for childhood diarrhea in Cameroon, also did not consider E. coli CFUs in the drinking water, but rather the type of water storage container, specifically the size of the mouth of the storage container and source of drinking water (Tambe, Nzefa, & Nicoline, 2015). Many studies compare treated drinking water to odds of developing diarrhea (Tarekegn & Enqueselassie, 2012). However, these do not look at water quality, nor evaluate the E. coli CFU in drinking water. In fact, one study, includes "Is microorganism cause of diarrhea" in the association variable table, without ever elaborating further what microorganisms were accounted for, or how they determined these did or did not cause diarrhea (Tarekegn & Enqueselassie, 2012).

Our study found that E. coli, as a determinate of water quality, was not a risk factor for childhood diarrhea. Though this may lead to conclusions that E. coli may not be the best variable to determine water quality, there is research that supports E. coli should continue to be used as an indicator for waterborne diarrhea risk (Ercumen et al., 2017; Gruber et al., 2014). E. coli is recommended to be used as a fecal indicator of waterborne diarrhea risk, as long as studies are carefully designed to address many potential biases (Ercumen et al., 2017; Gruber et al., 2014). However, even with the WHO recommendations of zero E. coli CFU/100 mL in drinking water, there is not a guarantee other pathogens are also absent, nor does it guarantee safety of the drinking water (Bain et al., 2021; WHO, 2017). Other indicators that could be used for water quality, which may present fewer challenges, include free residual chlorine levels, water source, turbidity of drinking water, bacterial spores or bacteriophages.

This study revealed that at baseline, there was an inverse association between animals (cows and chickens) observed in household compounds and caregiver reported diarrhea. Many studies examining determinants of childhood diarrhea do not include animal variables (Asfaha et al., 2018; Girma, Gobena, Medhin, Gasana, & Roba, 2018; Sinharoy et al., 2016; Tambe et al., 2015; Tarekegn & Enqueselassie, 2012). However, some researchers have studied the association between domestic animal husbandry (also referred to as domestic exposure to food-producing animals) and cases of diarrhea in humans. A systematic review and meta-analysis, published in 2014, found that 20 out of 29 studies (69%), reported a positive association between domestic animal exposure and diarrhea (Zambrano et al., 2014). This relationship is still rather understudied, and quality literature is still sparse, as noted in the 2014 systematic review; in fact, only three articles included in the review addressed E. coli. All three of these studies had a positive association between animal exposure and diarrhea (Zambrano et al., 2014). Another analysis, using data from 20 Sub-Saharan African countries, found varied results among the association between livestock ownership and diarrhea: 14 ORs greater than 1, 10 ORs less than one, and 6 null ORs (Kaur et al., 2017). In Rwanda, their findings had a OR=1 (95% CI= 0.93, 1.08) (Kaur et al., 2017). The pooled OR was null and the overall reported result was that household livestock ownership did not suggest a clear risk for childhood diarrhea (Kaur et al., 2017). A study in Uganda also had mixed results, this study looked more specifically at different animals owned and number of animals. They found that families with >5 poultry had higher rates of diarrhea in both adjusted and unadjusted models, looking at prevalence rate (PR); PR=1.83 (95% CI= 1.04, 3.23) and PR = 1.09 (95% CI= 0.58, 2.05) respectively. However, their findings for cows, sheep/goats were not consistent with the poultry finding and saw an inverse relationship in both adjusted and unadjusted models between animal ownership and diarrhea (Ercumen et al., 2020). Further research and attention towards the animal husbandry variable and association with diarrhea could be promising towards better understanding risk factors and protective factors for childhood diarrhea.

While water quality and observed animals in the compound were not significantly associated with caregiver reported diarrhea in our study, univariable regressions indicated a significant positive association between current breastfeeding status and caregiver reported diarrhea. Current breastfed status was much higher in children <24 months 495 (86.4%) as compared to children <60 months 103 (10.7%). This study had another variable, if a child had ever been breastfed. It was not significant in the univariable regression and descriptive characteristics showed 2 (0.3%) children <24 and 11 (1.1%) of children <60 months old had never been breastfed. Thus, an overwhelming majority of children <60 months had been breastfed in their lifetime, even though many children were not currently breastfeeding. The age group currently breastfeeding is younger, and diarrhea is more prevalent in the younger group across almost all reporting types (Table 3). Additionally, research supports that diarrhea is more prevalent across the younger age group <24 months (Canizalez-Roman et al., 2016; Nguyen, Le Van, Le Huy, & Weintraub, 2004). This study did not further investigate the relationship between breastfeeding and caregiver reported diarrhea. As our univariable regression showed a positive relationship between breastfeeding status and caregiver reported diarrhea, this differs from literature. Research has demonstrated that typically there is a protective effect of breastfeeding on childhood diarrhea (Arifeen et al., 2001; Lamberti, Fischer Walker, Noiman, Victora, & Black, 2011; McCormick et al., 2021; Richard et al., 2018). It is expected that a confounding assessment analysis would determine other confounders associated with this relationship and thus we recommend that future research incorporate a more thorough confounding assessment related to current breastfeeding status.

Additionally, our study did not find maternal/respondent schooling, child sex, respondent sex, or electricity significant in the univariable regression and thus these variables were not adjusted for in the multivariable regression. In other studies, maternal schooling, source of drinking water, child sex, birth order, and distance to water source were adjusted for (Asfaha et al., 2018; Sinharoy et al., 2016). Additionally, type of toilet facility and latrine ownership has been demonstrated to be associated with diarrhea morbidity incidence, and that was not adjusted for or included in our analysis (Tambe et al., 2015; Tarekegn & Enqueselassie, 2012). However, consistent with other studies, our multivariable model adjusted for poverty status, household size, and child's age in months (Sinharoy et al., 2016; Tambe et al., 2015).

Strengths and limitations

Our study has several strengths. First, we collected this data at the household level. As our data showed, the highest prevalence of diarrhea was reported by caregivers. Thus, by collecting data from caregivers at the household level, we may have captured incidents of diarrhea that would have been missed if we used data collected by community health workers, or at the health clinics. Furthermore, by collecting data ourselves, we have access to the raw data, data collectors, and the original survey questions and therefore are not limited in ways that national data or using data collected by others would pose. Another strength of our analysis, is that we clearly articulate which variables were controlled for in the multivariable analysis. Many similar studies do not clearly specify which variables were controlled for (Girma et al., 2018; Tambe et al., 2015). This makes this study clear, and replicable. Furthermore, our 7-day recall period, much shorter than the often used 2-week recall period, lessens recall bias.

One of the limitations for this study is that this analysis was restricted to baseline. Because of this, it is a "snapshot" of one data collection time point, which has known drawbacks. Seasonality is not factored into this analysis. Since we did not use all of the timepoints where data was collected, this is not a longitudinal analysis. Thus, when using only one time point, it is not possible to establish causal links. Thus, all data from the analysis cannot be discussed as causing or not causing the outcome of interest: caregiver reported diarrhea. Another limitation is that this study provides no prevention recommendations or treatment methods for diarrhea, nor does it consider other causes of diarrhea or diagnosing specific and contagious diarrheal diseases such as cholera or shigella. Due to limited capacity of the research team and time constraints, as previously mentioned, we did not include as many variables in our univariable regression. Analyzing more variables in our univariable model may have led to additional variables which should have been controlled for.

Public Health Implication and Recommendations

Results from this study have broader public health implications. These will be elaborated on, and specific recommendations will be given based on our findings.

A dose response for E. coli and diarrhea in adults is known (Barrag et al., 2021). There has been a call in many publications for a dose response to be further explored between E. coli and diarrhea for children under 24 months and under 60 months, as it still remains unknown (Barrag et al., 2021; Goddard, Ban, et al., 2020; Hodge et al., 2016). We also make this recommendation. However, in regards to our findings, and similar findings showing no association or minimal association between E. coli levels and water quality with diarrhea (Sinharoy et al., 2016), this dose response would also further determine if in fact an association exists, or further support our findings that there is no association.

Further, researchers and programmatic implementers should consider other factors, besides relying on E. coli as the indicator for water quality, or as a predictor of diarrhea. Water quality testing for E. coli levels is challenging, expensive, and time consuming. The findings in this study that there does not appear to be an association between water quality (as measured by E.

coli) and childhood diarrhea encourages other factors to be more heavily considered for predicting diarrhea. Diarrheagenic E. coli was only found in 37.36% of children under 60 months in dry season and 30.01% of children in Tanzania's rainy season (Vargas et al., 2004). Our study was only concerned with E. coli levels in drinking water; coupled with the Tanzania study, perhaps E. coli is not the best pathogen to be focused on. Thus, other pathogens in drinking water, such as norovirus, shigella, or parasites, should be further analyzed. As drinking water access, especially piped water and the availability of chlorine treatment, expands globally, perhaps E. coli and pathogen testing is becoming outdated, and less important than other water testing. Free residual chlorine, which when present allows researchers to determine none or minimal E. coli levels, may be a better approach in future studies. This is not only cheaper, but also more useful for understanding overall drinking water quality and cleanliness.

More diarrhea seems to be reported by caregivers, as noted in Table 3, when compared to community health worker and clinical diarrhea reported cases. This is consistent with literature that caregiver reported diarrhea is often used to measure diarrhea (Aiemjoy et al., 2018; Arnold et al., 2013). Therefore, it is recommended to continue using caregiver reported diarrhea at the household level for the most accurate child diarrhea cases. Clinic reports of diarrhea are unlikely to as accurately assess diarrhea for children under 24 months and under 60 months; and thus, these data should not be used for accurately assessing diarrhea prevalence. Community diarrhea assessments, through caregiver reported diarrhea, should be the recognized as the best approach for accurate childhood diarrhea cases and prevalence.

There should also be further research into animal ownership, and observed animals (cows and chickens) in household compounds and the association with childhood diarrhea. Our findings support this call for further research, and other literature recognizes the sparse evidence on this topic and also calls for further exploration and attention (Goddard, Ban, et al., 2020; Zambrano et al., 2014). This study found an inverse association. Several studies look at multiple variables in order to determine household socio-economic status and sociodemographic characteristics; these studies still did not consider animal ownership (Demissie, Yeshaw, Aleminew, & Akalu, 2021; Sinmegn Mihrete, Asres Alemie, & Shimeka Teferra, 2014). And, many studies use self-reported poverty status (Asfaha et al., 2018). The multivariable regression with the independent outcome variable being animals observed at baseline, was adjusted for poverty level in our study. This suggests that the argument households who can afford animals might be healthier and less likely to have diarrhea is not probable. As mentioned, published literature results vary when determining if and how animal ownership is associated with diarrhea. Conclusions differ between studies. Further researching this variable and its association to childhood diarrhea may find animals observed in the household compound to be associated with more frequent handwashing or higher respondent or caregiver education, variables that our study did not adjust for.

Conclusion

This study found when restricted to baseline, there was no association between drinking water quality and caregiver reported diarrhea among children in our study population. In addition, there was an inverse, but not statistically significant, relationship between chickens and cows observed in household compounds and caregiver reported diarrhea. Attention should be given to collecting similar observational data of animals in household compounds for future studies and interventions. We recommend that drinking water quality should not be the main predicting factor for childhood diarrhea; in fact, funds allocated to the time consuming and expensive drinking water analyses should be used for other aspects of studies or interventions.

Bibliography

Aiemjoy, K., Aragie, S., Gebresillasie, S., Fry, D. M., Dagnew, A., Hailu, D., . . . Keenan, J. D. (2018). Defining Diarrhea: A Population-Based Validation Study of Caregiver-Reported Stool Consistency in the Amhara Region of Ethiopia. *The American Journal of Tropical Medicine and Hygiene, 98*(4), 1013-1020. doi:10.4269/ajtmh.17-0806

Arifeen, S., Black, R. E., Antelman, G., Baqui, A., Caulfield, L., & Becker, S. (2001). Exclusive Breastfeeding Reduces Acute Respiratory Infection and Diarrhea Deaths Among Infants in Dhaka Slums. *Pediatrics, 108*(4), e67-e67. doi:10.1542/peds.108.4.e67

Arnold, B. F., Galiani, S., Ram, P. K., Hubbard, A. E., Briceño, B., Gertler, P. J., & Colford, J. M., Jr. (2013). Optimal recall period for caregiver-reported illness in risk factor and intervention studies: a multicountry study. *Am J Epidemiol, 177*(4), 361-370. doi:10.1093/aje/kws281

Asfaha, K. F., Tesfamichael, F. A., Fisseha, G. K., Misgina, K. H., Weldu, M. G., Welehaweria, N. B., & Gebregiorgis, Y. S. (2018). Determinants of childhood diarrhea in Medebay Zana District, Northwest Tigray, Ethiopia: a community based unmatched case–control study. *BMC Pediatrics, 18*(1), 120. doi:10.1186/s12887-018-1098-7

Bain, R., Johnston, R., Khan, S., Hancioglu, A., & Slaymaker, T. (2021). Monitoring Drinking Water Quality in Nationally Representative Household Surveys in Low- and Middle-Income Countries: Cross-Sectional Analysis of 27 Multiple Indicator Cluster Surveys 2014–2020. *Environmental Health Perspectives, 129*(9), 097010. doi:doi:10.1289/EHP8459

Barrag, J. L. M., Cuesta, L. D. I., & Susa, M. S. R. (2021). Quantitative microbial risk assessment to estimate the public health risk from exposure to enterotoxigenic E. coli in drinking water in the rural area of Villapinzon, Colombia. *Microbial Risk Analysis, 18*. doi:10.1016/j.mran.2021.100173

Canizalez-Roman, A., Flores-Villaseñor, H. M., Gonzalez-Nuñez, E., Velazquez-Roman, J., Vidal, J. E., Muro-Amador, S., . . . León-Sicairos, N. (2016). Surveillance of Diarrheagenic Escherichia coli Strains Isolated from Diarrhea Cases from Children, Adults and Elderly at Northwest of Mexico. *Frontiers in Microbiology, 7*. doi:10.3389/fmicb.2016.01924

CDC. (2018). Diarrhea: Common Illness, Global Killer. Retrieved from https://www.cdc.gov/healthywater/pdf/global/programs/globaldiarrhea508c.pdf

Clasen Research Group - Emory University Rollins School of Public Health. (n.d.). Retrieved from https://scholarblogs.emory.edu/clasenresearch/?page_id=16

Clasen, T., Schmidt, W.-P., Rabie, T., Roberts, I., & Cairncross, S. (2007). Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *BMJ*. doi:10.1136/bmj.39118.489931.BE

Cook, A., & Sheikh, A. (2000). Descriptive statistics (Part 2): Interpreting study results. *Primary Care Respiratory Journal, 8*(1), 16-17. doi:10.1038/pcrj.2000.6

Demissie, G. D., Yeshaw, Y., Aleminew, W., & Akalu, Y. (2021). Diarrhea and associated factors among under five children in sub-Saharan Africa: Evidence from demographic and health surveys of 34 sub-Saharan countries. *PLoS One, 16*(9), e0257522-e0257522. doi:10.1371/journal.pone.0257522

Dolstad, H. A., Franke, M. F., Vissieres, K., Jerome, J. G., Ternier, R., & Ivers, L. C. (2021). Factors associated with diarrheal disease among children aged 1-5 years in a cholera epidemic in rural Haiti. *PLoS Negl Trop Dis, 15*(10), e0009726. doi:10.1371/journal.pntd.0009726

Dufour, A. (1977). *Escherichia coli: the fecal coliform*: ASTM International.

Enger, K. S., Nelson, K. L., Clasen, T., Rose, J. B., & Eisenberg, J. N. S. (2012). Linking Quantitative Microbial Risk Assessment and Epidemiological Data: Informing Safe Drinking Water Trials in Developing Countries. *Environmental Science & Technology, 46*(9), 5160-5167. doi:10.1021/es204381e

Ercumen, A., Arnold, B. F., Naser, A. M., Unicomb, L., Colford Jr, J. M., & Luby, S. P. (2017). Potential sources of bias in the use of Escherichia coli to measure waterborne diarrhoea risk in low-income settings. *Tropical Medicine & International Health, 22*(1), 2-11. doi:https://doi.org/10.1111/tmi.12803

Ercumen, A., Prottas, C., Harris, A., Dioguardi, A., Dowd, G., & Guiteras, R. (2020). Poultry Ownership Associated with Increased Risk of Child Diarrhea: Cross-Sectional Evidence from Uganda. *The American Journal of Tropical Medicine and Hygiene, 102*(3), 526-533. doi:10.4269/ajtmh.19-0012

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., & Colford, J. M., Jr. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infect Dis, 5*(1), 42-52. doi:10.1016/s1473-3099(04)01253-8

Girma, M., Gobena, T., Medhin, G., Gasana, J., & Roba, K. T. (2018). Determinants of childhood diarrhea in West Gojjam, Northwest Ethiopia: a case control study. *The Pan African medical journal, 30*, 234-234. doi:10.11604/pamj.2018.30.234.14109

Goddard, F. G. B., Ban, R., Barr, D. B., Brown, J., Cannon, J., Colford, J. M., . . . Clasen, T. (2020). Measuring Environmental Exposure to Enteric Pathogens in Low-Income Settings: Review and Recommendations of an Interdisciplinary Working Group. *Environmental Science & Technology, 54*(19), 11673-11691. doi:10.1021/acs.est.0c02421

Goddard, F. G. B., Chang, H. H., Clasen, T. F., & Sarnat, J. A. (2020). Exposure measurement error and the characterization of child exposure to fecal contamination in drinking water. *NPJ Clean Water, 3*(1), 1-9. doi:10.1038/s41545-020-0063-9

Grace-Martin, K. (2015). What is a Logit Function and Why Use Logistic Regression? *The Analysis Factor.* Retrieved from https://www.theanalysisfactor.com/what-is-logit-function/

Gruber, J. S., Ercumen, A., & Colford, J. M., Jr. (2014). Coliform Bacteria as Indicators of Diarrheal Risk in Household Drinking Water: Systematic Review and Meta-Analysis. *PLoS One, 9*(9), e107429. doi:10.1371/journal.pone.0107429

Hodge, J., Chang, H. H., Boisson, S., Collin, S. M., Peletz, R., & Clasen, T. (2016). Assessing the Association between Thermotolerant Coliforms in Drinking Water and Diarrhea: An Analysis of Individual–Level Data from Multiple Studies. *Environmental Health Perspectives, 124*(10), 1560-1567. doi:doi:10.1289/EHP156

Howard, G., Pedley, S., & Tibatemwa, S. (2006). Quantitative microbial risk assessment to estimate health risks attributable to water supply: can the technique be applied in developing countries with limited data? *J Water Health, 4*(1), 49-65.

IBM. (2022, 2/23/22). What is Logistic regression? Retrieved from https://www.ibm.com/topics/logistic-regression

Karambizi, N. U., McMahan, C. S., Blue, C. N., & Temesvari, L. A. (2021). Global estimated Disability-Adjusted Life-Years (DALYs) of diarrheal diseases: A systematic analysis of data from 28 years of the global burden of disease study. *PLoS One, 16*(10), e0259077. doi:10.1371/journal.pone.0259077

Kaur, M., Graham, J. P., & Eisenberg, J. N. S. (2017). Livestock Ownership Among Rural Households and Child Morbidity and Mortality: An Analysis of Demographic Health Survey Data from 30 Sub-Saharan African Countries (2005-2015). *The American Journal of Tropical Medicine and Hygiene, 96*(3), 741-748. doi:10.4269/ajtmh.16-0664

Kirby, M. A., Nagel, C. L., Rosa, G., Iyakaremye, L., Zambrano, L. D., & Clasen, T. F. (2016). Faecal contamination of household drinking water in Rwanda: A national cross-sectional study. *Science of The Total Environment, 571*, 426-434. doi:https://doi.org/10.1016/j.scitotenv.2016.06.226

Lamberti, L. M., Fischer Walker, C. L., Noiman, A., Victora, C., & Black, R. E. (2011). Breastfeeding and the risk for diarrhea morbidity and mortality. *BMC Public Health, 11*(3), S15. doi:10.1186/1471-2458-11-S3-S15

LaValley, M. P. (2008). Logistic Regression. *Circulation, 117*(18), 2395-2399. doi:doi:10.1161/CIRCULATIONAHA.106.682658

Lim, M. L., & Wallace, M. R. (2004). Infectious diarrhea in history. *Infectious Disease Clinics of North America, 18*(2), 261-274. doi:https://doi.org/10.1016/j.idc.2004.01.006

Luby, S. P., Halder, A. K., Huda, T. M., Unicomb, L., Islam, M. S., Arnold, B. F., & Johnston, R. B. (2015). Microbiological Contamination of Drinking Water Associated with Subsequent Child Diarrhea. *The American Society of Tropical Medicine and Hygiene, 93*(5), 904-911. doi:10.4269/ajtmh.15-0274

McCormick, B. J. J., Richard, S. A., Murray-Kolb, L. E., Kang, G., Lima, A. A. M., Mduma, E., . . . Investigators, M.-E. N. (2021). Full breastfeeding protection against common enteric bacteria and viruses: results from the MAL-ED cohort study. *The American Journal of Clinical Nutrition, 115*(3), 759-769. doi:10.1093/ajcn/nqab391

McMahan, Z. H., & DuPont, H. L. (2007). Review article: the history of acute infectious diarrhoea management--from poorly focused empiricism to fluid therapy and modern pharmacotherapy. *Aliment Pharmacol Ther, 25*(7), 759-769. doi:10.1111/j.1365-2036.2007.03261.x

Nguyen, T. V., Le Van, P., Le Huy, C., & Weintraub, A. (2004). Diarrhea caused by rotavirus in children less than 5 years of age in Hanoi, Vietnam. *Journal of clinical microbiology, 42*(12), 5745-5750. doi:10.1128/JCM.42.12.5745-5750.2004

Onda, K., LoBuglio, J., & Bartram, J. (2012). Global Access to Safe Water: Accounting for Water Quality and the Resulting Impact on MDG Progress. *International Journal of Environmental Research and Public Health, 9*(3), 880-894. Retrieved from https://www.mdpi.com/1660-4601/9/3/880

Organization, W. H. (1997). Guidelines for drinking-water quality Second Edition. *Volume 3: Surveillance and control of community supplies*. Retrieved from https://www.who.int/water_sanitation_health/dwq/2edvol3a.pdf

Pickering, A. J., Ercumen, A., Arnold, B. F., Kwong, L. H., Parvez, S. M., Alam, M., . . . Luby, S. P. (2018). Fecal Indicator Bacteria along Multiple Environmental Transmission Pathways (Water, Hands, Food, Soil, Flies) and Subsequent Child Diarrhea in Rural Bangladesh. *Environmental Science & Technology, 52*(14), 7928-7936. doi:10.1021/acs.est.8b00928

Prevention, C. f. D. C. a. (2022). E. coli (Escherichia coli). Retrieved from https://www.cdc.gov/ecoli/index.html

Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M. C., . . . Johnston, R. (2019). Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *International Journal of Hygiene and Environmental Health, 222*(5), 765-777. doi:https://doi.org/10.1016/j.ijheh.2019.05.004

Qadri, F., Svennerholm, A.-M., Faruque, A. S. G., & Sack, R. B. (2005). Enterotoxigenic <i>Escherichia coli</i> in Developing Countries: Epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. *Clinical Microbiology Reviews, 18*(3), 465-483. doi:doi:10.1128/CMR.18.3.465-483.2005

Ranganathan, P., Aggarwal, R., & Pramesh, C. S. (2015). Common pitfalls in statistical analysis: Odds versus risk. *Perspectives in clinical research, 6*(4), 222-224. doi:10.4103/2229-3485.167092

Richard, S. A., McCormick, B. J. J., Seidman, J. C., Rasmussen, Z., Kosek, M. N., Rogawski, E. T., . . . On Behalf Of The Mal-Ed Network, I. (2018). Relationships among Common Illness Symptoms and the Protective Effect of Breastfeeding in Early Childhood in MAL-ED: An Eight-Country Cohort Study. *The American Journal of Tropical Medicine and Hygiene, 98*(3), 904-912. doi:10.4269/ajtmh.17-0457

Rosa, G., Huaylinos, M. L., Gil, A., Lanata, C., & Clasen, T. (2014). Assessing the consistency and microbiological effectiveness of household water treatment practices by urban and rural populations claiming to treat their water at home: a case study in Peru. *PLoS One, 9*(12), e114997. doi:10.1371/journal.pone.0114997

Rosa, G., Kelly, P., & Clasen, T. (2016). Consistency of Use and Effectiveness of Household Water Treatment Practices Among Urban and Rural Populations Claiming to Treat Their Drinking Water at Home: A Case Study in Zambia. *Am J Trop Med Hyg, 94*(2), 445-455. doi:10.4269/ajtmh.15-0563

Sinharoy, S. S., Schmidt, W.-P., Cox, K., Clemence, Z., Mfura, L., Wendt, R., . . . Clasen, T. (2016). Child diarrhoea and nutritional status in rural Rwanda: a cross-sectional study to explore contributing environmental and demographic factors. *Tropical Medicine & International Health, 21*(8), 956-964. doi:https://doi.org/10.1111/tmi.12725

Sinmegn Mihrete, T., Asres Alemie, G., & Shimeka Teferra, A. (2014). Determinants of childhood diarrhea among underfive children in Benishangul Gumuz Regional State, North West Ethiopia. *BMC Pediatrics, 14*(1), 102. doi:10.1186/1471-2431-14-102

Tambe, A. B., Nzefa, L. D., & Nicoline, N. A. (2015). Childhood Diarrhea Determinants in Sub-Saharan Africa: A Cross Sectional Study of Tiko-Cameroon. *Challenges, 6*(2), 229-243. Retrieved from https://www.mdpi.com/2078-1547/6/2/229

Tarekegn, M., & Enqueselassie, F. (2012). A case control study on determinants of diarrheal morbidity among under-five children in Wolaita Soddo Town, Southern Ethiopia. *Ethiopian Journal of Health Development, 26*, 78-85.

Troeger, C., Blacker, B. F., Khalil, I. A., Rao, P. C., Cao, S., Zimsen, S. R. M., . . . Reiner, R. C., Jr. (2018). Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Infectious Diseases, 18*(11), 1211-1228. doi:10.1016/S1473-3099(18)30362-1

Vargas, M., GascóN, J., Casals, C., Schellenberg, D., Urassa, H., Kahigwa, E., . . . Vila, J. (2004). ETIOLOGY OF DIARRHEA IN CHILDREN LESS THAN FIVE YEARS OF AGE IN IFAKARA, TANZANIA. *The American Journal of Tropical Medicine and Hygiene Am J Trop Med Hyg, 70*(5), 536-539. doi:10.4269/ajtmh.2004.70.536

Wagner, E. G., Lanoix, J. N., & World Health, O. (1958). Excreta disposal for rural areas and small communities / Edmund G. Wagner, J. N. Lanoix. In. Geneva: World Health Organization.

Whalley, B. (n.d.). 25.1 Link Functions. *Just enough R.* Retrieved from https://benwhalley.github.io/just-enough-r/link-functions.html

WHO. (2017). WHO Guidelines Approved by the Guidelines Review Committee. In *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum*. Geneva: World Health Organization

Copyright © World Health Organization 2017.

Wolf, J., Prüss-Ustün, A., Cumming, O., Bartram, J., Bonjour, S., Cairncross, S., . . . Higgins, J. P. (2014). Assessing the impact of drinking water and sanitation on diarrhoeal disease in low- and middle-income settings: systematic review and meta-regression. *Trop Med Int Health, 19*(8), 928-942. doi:10.1111/tmi.12331

World Health Oganization, U. (2009). Diarrhoea: Why children are still dying and what can be done. 69.

Zambrano, L. D., Levy, K., Menezes, N. P., & Freeman, M. C. (2014). Human diarrhea infections associated with domestic animal husbandry: a systematic review and meta-analysis. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 108*(6), 313-325. doi:10.1093/trstmh/tru056

Appendix

Variable Code	Variable Description	Data Type	Household or Individual	Collection Type
diarrhea_caregiver	In the last 7 days, child reported to have diarrhea by caregiver's definition	Categorical: 0=No 1=Yes	Individual	Survey
diarrhea_CHW_clinic	In the last 7 days, child reported to be taken to CHW or clinic for diarrhea treatment	Categorical: 0=No 1=Yes	Individual	Survey
diarrheaWHO	In the last 7 days, child reported to have 3 or more loose stools in 24 hours	Categorical: 0=No 1=Yes	Individual	Survey
diarrheaCHW	In the last 7 days, child reported to be taken to CHW for diarrhea treatment	Categorical: 0=No 1=Yes	Individual	Survey
diarrheaclinic	In the last 7 days, child reported to be taken to clinic for diarrhea treatment	Categorical: 0=No 1=Yes	Individual	Survey
type1	In the last 7 days, child reported to pass Type 1 stool	Categorical: 0=No 1=Yes 99= Don't Know	Individual	Survey
type2	In the last 7 days, child reported to pass Type 2 stool	Categorical: 0=No 1=Yes	Individual	Survey

		99= Don’t Know		
type3	In the last 7 days, child reported to pass Type 3 stool	Categorical: 0=No 1=Yes 99= Don’t Know	Individual	Survey
type4	In the last 7 days, child reported to pass Type 4 stool	Categorical: 0=No 1=Yes 99= Don’t Know	Individual	Survey
type5	In the last 7 days, child reported to pass Type 5 stool	Categorical: 0=No 1=Yes 99= Don’t Know	Individual	Survey
type45_freq	Reported max number of times child passed 4 or 5 type stool in 24 hrs in past 7 days	Categorical: 0=No 1=Yes 99= Don’t Know	Individual	Survey
CatWQ2	Number of E.coli Colony Forming Units per 100 mL broken down into categories	Categorical: <1, 1-10, 11-100, 100+	Household	Water Sample & Analysis
WQ	Number of E.coli Colony Forming Units per 100 mL	Continuous	Household	Water Sample & Analysis
risk	WHO Water Quality Risk Levels	Categorical: 0=<2 CFU/100mL (no detectable E.coli) 1=1-10 CFU/100mL 2=11-100 CFU/100mL	Household	Water Sample & Analysis

		3=>100 CFU/100mL		
sex	Sex of the Child	Categorical: 1=Male 2=Female 3=Other	Individual	Survey
res_primedu	Respondent Education Reported at Baseline	Categorical: 0=Never attended school 1=Some preschool or primary school 2=Primary school complete or higher	Household	Survey
ubu12.x	Household Reported to Belong to Lowest Two Categories of Government-defined SES	Categorical: 1=Ubu 1 or 2 2=Not Ubu 1 or 2	Household	Survey
agemonths	Age in Months of the Child	Continuous	Individual	Survey
hhsize	Household Size Reported at Baseline	Continuous	Household	Survey
electricity	Access to Electricity Reported at Baseline	Categorical: 0=No 1=Yes	Household	Survey
animalcomp	Observed Chickens/Cows or Evidence of Chickens/Cows in Compound at Baseline	Categorical: 0=No 1=Yes	Household	Observation
respon_sex	Sex of Survey Respondent	Categorical: 1=Male 2=Female	Household	Survey

		3=Other		
breastfedstatus	Child reported to be currently breastfed	Categorical: 0=No 1=Yes 99= Don’t Know	Individual	Survey
res_age	Age of Respondent at Baseline	Continuous	Household	Survey

Appendix Table 1: Variable Descriptions