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:

_____ _____

Laura Divens Zambrano Date

Predictors of diarrhea and respiratory disease and use of serological markers to assess the health impact of a household filtration intervention among young children in Western Province, Rwanda

By

Laura Divens Zambrano

Doctor of Philosophy

Environmental Health Sciences

_____

Thomas Clasen, Ph.D., M.Sc., J.D.
Advisor

_____

Jeremy Sarnat, Sc.D., M.Sc., M.S.
Committee Member

_____

Dana Flanders, D.Sc., M.P.H., M.D., M.A.
Committee Member

_____

Dana Barr, Ph.D.
Committee Member

_____

Corey Nagel, Ph.D., M.P.H., R.N.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D. Dean of the James T. Laney School of Graduate Studies

_____

Date

Predictors of diarrhea and respiratory disease and use of serological markers to assess the health impact of a household filtration intervention among young children in Western Province, Rwanda

By

Laura Divens Zambrano

M.P.H., The George Washington University, 2010

B.A., St. Mary's College of Maryland, 2006

Advisor: Thomas Clasen, Ph.D., M.Sc., J.D.

An abstract of

a dissertation submitted to the

Faculty of the James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Environmental Health Sciences

2016

Abstract

Predictors of diarrhea and respiratory disease and use of serological markers

to assess the health impact of a household filtration intervention among

young children in Western Province, Rwanda

By Laura Divens Zambrano

Diarrhea and acute lower respiratory infection (ALRI) are the two largest contributors to childhood morbidity and mortality in Sub-Saharan Africa and are primarily attributable to poor water quality, inadequate sanitation and hygiene and exposure to household air pollution from cooking with biomass fuels. Household filtration interventions significantly reduce diarrheal disease, even in the absence of any improvements to sanitation facilities. Interventions addressing household air pollution exposures have had varying degrees of success given the substantial reductions in fine particulate matter exposure (e.g., $PM_{2.5}$) required to have a measurable health effect. This dissertation leverages the baseline round of a large cluster-randomized controlled trial of a household water filter and a wood-burning rocket stove to characterize household characteristics and environmental exposures that are associated with diarrhea and ALRI. Seroconversion between baseline and follow-up against various enteropathogens was also assessed relative to study arm assignment and recent diarrheal disease.

Multilevel analyses accounting for the complex survey design of the two cross-sectional studies of the baseline data revealed several household characteristics that were associated with diarrhea and ALRI. Compared to protected spring water sources, piped water sources and dug wells were protective against diarrheal disease while standpipes, boreholes and surface water sources were associated with excess risk. Compared with pit latrines, diarrhea prevalence was higher in households that had shared sanitation and non-shared household composting toilets. Drinking water quality was not associated with diarrheal disease. While ALRI was not affected by household stove and cooking characteristics, $PM_{2.5}$ exposures varied significantly by cooking location, fuel use and stove type. The third study examining seroconversion against various enteropathogens found that the water filter intervention significantly decreased *Cryptosporidium* seroconversion, that both *Giardia* and *Cryptosporidium* were associated with recent diarrheal disease and that seroconversion peaked after 12 months of age.

This dissertation outlines particular household exposures that are disproportionately linked to childhood diarrhea and respiratory diseases, which should be useful to organizations and stakeholders working in this field. It also generates future directions for this research, particularly in the application of serological markers to the evaluation of water and sanitation interventions.

Predictors of diarrhea and respiratory disease and use of serological markers to assess the health impact of a household filtration intervention among young children in Western Province, Rwanda

By

Laura Divens Zambrano

M.P.H., The George Washington University, 2010

B.A., St. Mary's College of Maryland, 2006

Advisor: Thomas Clasen, Ph.D., M.Sc., J.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Environmental Health Sciences

2016

Acknowledgements

Many people were involved in making my time here at Emory a success. I feel indebted to my five committee members, who provided immense technical help and moral support along the way: Thomas Clasen, my committee chair, for his incredible mentorship and friendship throughout my time at Emory and in Rwanda, whose sense of humor eased me through the hardest times; Corey Nagel, for his unwavering patience through all phases of this project, and for his companionship during our travels abroad; Jeremy Sarnat, for providing great leadership to our program and for his steadfast support of my professional and academic pursuits, which made my involvement in this project possible; Dana Flanders, for his thoughtful advice that challenged me to grow as an environmental epidemiologist; and to Dana Barr, for her past and continued support on many of the biological measurements and exposure assessments involved with this project.

Thank you to our study enumerators, who not only made this project possible on the ground but became my in-country family: Florien, Media, Ephrem, Gloriose, Laurien, Monique, Bernard, Jonathan, Amon, Vianney, Maurice, Cyridion, Gorette, Charles, Ange, Emmy, Mado, Lambert and Francine. Their dedication to this project truly inspired me.

My colleagues at the London School of Hygiene and Tropical Medicine played critical roles in both the design and implementation of the trial and the analysis of its forthcoming results. I am grateful to Ghislaine Abadie Rosa and Miles Kirby for all they have done to make this trial possible and for their friendship.

Thank you to my colleagues at the Rwanda National Reference Laboratory, Dr. Emil Ivan and Dr. John Rusine, for their logistical and technical support of this project. Thank you to Dr. Jeffrey Priest and Dr. Patrick Lammie at the Centers for Disease Control and Prevention, for opening up their laboratory to me for the dried blood spot analysis and for providing me with superb technical and scientific direction.

Thank you to the faculty and staff of the Rollins School of Public Health. I have gained so much from your mentorship and experience, in particular: Dr. Matthew Freeman, Dr. Karen Levy and Dr. Paige Tolbert.

Finally, thank you to my husband Daniel Zambrano and to my family and friends for their unwavering support and encouragement.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALRI - Acute Lower Respiratory Infection
ARI - Acute Respiratory Infection (Upper)
CDC - Centers for Disease Control and Prevention
CFU - Colony-Forming Unit
CHW - Community Health Worker
CO - Carbon monoxide
COHb - Carboxyhemoglobin
Cp17 - *Cryptosporidium* 17kDa protein
Cp23 - *Cryptosporidium* 23 kDa protein
Cp27 - *Cryptosporidium* 27 kDa protein
CRP - C-Reactive Protein
CtxB - β subunit of cholera enterotoxin
DALY - Disability-Adjusted Life-Years
DBS - Dried Blood Spot
DHS - Demographic and Health Survey
EED - Environmental Enteric Dysfunction
EHO - Environmental Health Officer
ELISA - Enzyme-Linked Immunosorbent Assay
ETEC - Enterotoxigenic *E. coli*
EtxB - β subunit of *E. coli* heat-labile enterotoxin
FIB - Fecal Indicator Bacteria
GBD - Global Burden of Disease
GST - Glutathione-S-transferase
HAP - House Air Pollution
HPEM -Harvard Personal Exposure Monitor
HWT - Household Water Treatment
IgG - Immunoglobulin G
IL - Interleukin
IMCI - Integrated Management of Childhood Illness
JMP - Joint Monitoring Programme
LecA - Lectin Adhesion Antigen
LMICs - Low/Middle Income Countries
LPG - Liquid Petroleum Gas
LPS - Lipopolysaccharide
MDGs - Millennium Development Goals
MUAC - Middle-Upper Arm Circumference
NRL - Rwandan National Reference Laboratory
PCA - Principal Components Analysis
PM - Particulate Matter
POU - Point-Of-Use
ROC - Receiver Operator Characteristic
SAPE - Streptavidin-Phycoerythrin
SpO_2 - Arterial Oxygen Saturation
TNF - Tumor Necrosis Factor
TNTC - Too numerous to count
TTC - Thermotolerant Coliform Counts
VSP - Variant-Specific Protein
WASH -Water, Sanitation and Hygiene
WHO - World Health Organization
WHZ - Weight-for-Height Z scores

Chapter 1 – Introduction

Acute lower respiratory infection (ALRI) and diarrheal disease are leading causes of morbidity and mortality among children under-5 years old in Sub-Saharan Africa. According to the 2010 Global Burden of Disease (GBD) study, ALRI and diarrhea are responsible for 4.6% and 3.6% of global disease burden, respectively [1]. Household air pollution (HAP) due to inefficient biomass-burning cookstoves alone was associated 4.3% of all global DALYs in 2010, making it the 3rd leading risk factor for global burden of disease [2]. HAP is the primary contributor of respiratory disease in children under 5 and a leading cause of chronic bronchitis and chronic obstructive pulmonary disease (COPD) among women [2,3]. Among the top non-communicable contributors to global DALYs, HAP is the only exposure that disproportionately affects children [2]. In addition to the direct link between HAP and development of lower respiratory infections (LRIs), HAP has been associated with inflammatory processes linked to cardiovascular sequelae. Today, diarrheal disease still accounts for approximately 1.5 million deaths annually [4], with many of these deaths occurring among young children in Sub-Saharan Africa [2].

Despite the fact that Rwanda is driving the most pronounced reduction of childhood mortality of any other country in the world, lack of access to improved water sources contributes substantially to diarrheal disease mortality in Rwanda and in the region as a whole [2]. In order to address this disease burden—and to reduce the costs and environmental burden associated with cooking fuels—DelAgua Health Rwanda (DelAgua), under the authority of the Ministry of Health, Rwanda, will distribute advanced water filters and high efficiency cook stoves to the poorest (*Ubudehe* levels 1 and 2)

households in Western Province, Rwanda. The intervention is being rolled out in three phases. Phase 3 envisions distribution throughout the entire country, subject to the results from the first two phases.

Phase 1, which was completed in late 2012 and early 2013, entailed a pilot distribution of the intervention filters and cookstoves by DelAgua to approximately 2200 households in 15 villages in the western portion of the country [5]. An evaluation used a parallel household-randomized controlled trial design of three rural villages in order to assess uptake and use of the intervention and the impact of the intervention on microbial water quality and air quality near household cooking areas [6]. This phase yielded household-reported uptake of water filters of 89.2% and a significant reduction of thermotolerant coliforms (TTC) (an indicator of fecal contamination) in household water samples; 86.8% (95% CI: 84.9-88.6%) of water samples from intervention households were TTC-free vs. 22.4% (95% CI: 20.1-24.6%) of control households [6]. Reported and observed exclusive use of the intervention stove was more inconsistent, but still yielded a 48% ($p<0.005$) reduction in area $PM_{2.5}$ concentrations around the cooking area in intervention vs. control households [6].

Phase 2 of the intervention involves distribution of the filters and stoves to the poorest 30% of the population of Rwanda's Western Province. While the evaluation includes assessments of uptake and impacts on exposure, the main focus is on health effects. The design is a cluster-randomized controlled trial in which the intervention is randomly assigned by administrative sector. Sectors were chosen as the unit of randomization since

they comprise the catchment area for health clinics. The research team randomized 72 sectors (consisting of about 100,000 households) to receive the intervention first, while the control arm, consisting of 24 sectors (about 40,000 households) will receive the intervention after the study period (Figure 1.1). In this main "sector-level" study, the impact of the intervention will be assessed using sector-level clinic data and community health worker (CHW) records. Nested within this overall sector-level study is a more in-depth assessment of 174 randomly selected villages (the "village-level sub-study") (Figure 1.2). This involves air and water sampling together with more extensive exposure and health collected from up to 10 households in each village. Following enrollment and a baseline assessment from August through December 2014, each participating household and will undergo follow-up visits on a quarterly basis throughout 2015. Figure 1 shows a schematic of the evaluation of Phase 2.

A sub-population within the village-level study (2 households per village) will be selected for the biomarker study. Dried blood spots (DBS) samples will be collected from household cooks and from children under-5 to assess biomarkers of inflammation, which have been linked to both exposures and ALRI disease. In addition, DBS disc samples containing a smaller volume of blood will be collected from children 6 to 12 months old at baseline and again from these same children at 6 months (Round 2 of follow-up) to analyze serum for seroconversion against common enteric pathogens.

This proposed use of DBS to obtain objective and quantitative biological measurements of measurable physiological effects could be a potential solution to the subjective nature of

disease outcome ascertainment, a particular problem in assessing the impact of environmental interventions that cannot be blinded. Biomarkers offer the potential for improving the consistency of diagnoses of these diseases associated with ALRI and environmental enteropathy and can also directly reflect exposure to contaminated air and drinking water. Biomarkers may be able to inform and improve the reliability of standardized disease outcome classifications. Recognizing the potential contribution of biomarker data, there are increasing calls for such data to be incorporated into future clean cookstove trials [7,8]. More extensive biomarker research is forthcoming. Analysis of seroconversion against enteric pathogens in young children can serve as a proxy for exposure to enteric pathogens among other household members.

The overall objective of my dissertation research is to assess household and environmental predictors of pneumonia and diarrhea in children under 5 years-old. We will also assess the relationship between intervention group assignment and serological markers of exposures to enteric pathogens. These activities will be undertaken in the context of a larger study in which I supported the collection and analysis of baseline data and data collection for other aspects of the Phase 2 evaluation.

Aims of Dissertation

Aim 1) To describe baseline activities and study design (principally related to water quality and diarrhea ascertainment), to perform a thorough assessment of baseline descriptive statistics of our study population and to determine factors associated with diarrhea point prevalence and 7-day period prevalence in children under-5 years-old prior to intervention distribution.

Aim 2) To describe baseline activities and study design (principally related to air quality and ALRI ascertainment), to perform a thorough assessment of baseline descriptive statistics of our study population and to determine factors associated with respiratory disease and ALRI point prevalence and 7-day prevalence in children under-5 years-old prior to intervention distribution.

Aim 3) To characterize serological markers of exposure to enteric pathogens and to apply these objective measures to assess the impact of point-of-use water filtration on diarrheal disease among children under-5 years-old in *Ubudehe* 1 and 2 households in Western Province.

These three aims will be addressed directly later in this dissertation. Chapters 2 and 3 will provide background on diarrheal disease and upper and lower respiratory disease among children in Sub-Saharan Africa, and interventions that have been evaluated to ameliorate these conditions. Chapter 4 will describe unifying methods used in the cross-sectional analyses of Aims 1 and 2, including our approaches to principal components analysis,

weighting and modeling approaches. Chapter 5 through 7 entail the manuscripts prepared to describe the methods and results after investigating the questions posed from the three dissertation aims. Chapter 8 concludes the dissertation with a summary of our findings, limitations and future implications of this research.

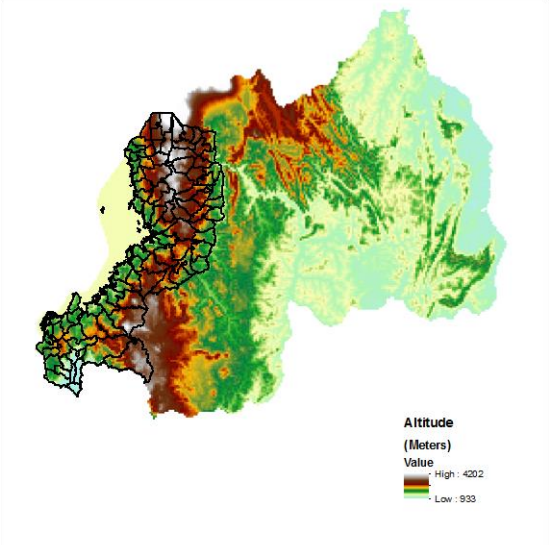


Figure 1.1a: Rwanda is a small mountainous country in East Africa. Western Province, demarcated by sector borders, lies in the highest region of the country.

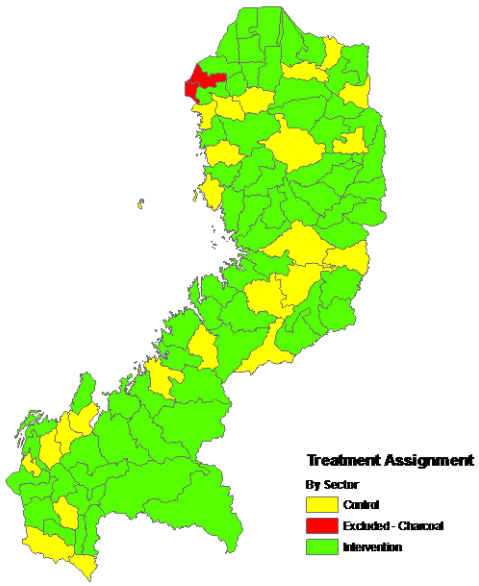


Figure 1.1b: Western Province is divided into 96 administrative sectors. The paired intervention in this study was randomized at a 3:1 ratio by sector, which also represent health clinic catchment areas.

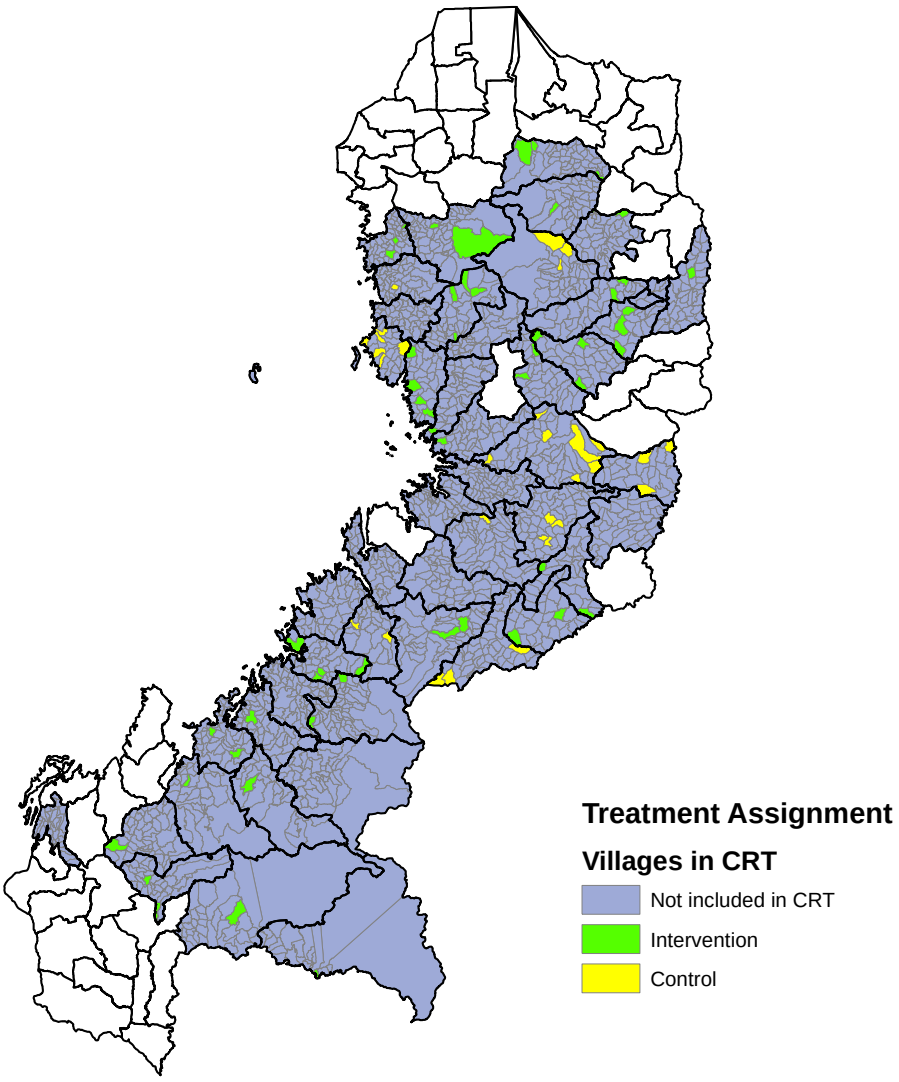


Figure 1.2: Location of study village clusters. 174 village clusters were selected through population proportional selection at a 1:1 ratio between intervention and control arms.

Chapter 2 – Diarrheal Disease and WASH

Diarrheal disease is among the leading causes of morbidity and mortality among children under-5 years-old in Sub-Saharan Africa and is responsible 3.6% of global disease burden [1]. Diarrheal disease is the 4th largest contributor to years of lives lost in eastern Sub-Saharan Africa, respectively, behind only HIV/AIDS and malaria [1]. Lack of access to safe drinking water is associated with about 116,000 deaths annually and nearly 8000 DALYs, as of 2010 estimates [2]. While this represents a significant decrease in mortality and morbidity attributable to unimproved water sources from 1990, many of the poorest households in Rwanda still lack access. Today, diarrheal disease still accounts for approximately 1.5 million deaths annually [4], with many of these deaths occurring among young children in Sub-Saharan Africa [2]. In addition to being linked to acute enteric infections, sustained contact with unsafe drinking water is also associated with growth stunting and undernutrition, which affects 20% of children in developing countries by causing intestinal inflammation and altering intestinal barrier and absorptive function [9]. Growth stunting and severe wasting are associated with 2.2 million deaths and 21% of DALYs among children under-5 in developing worldwide; the manifestations of stunting and wasting are evident at age 1 and may be considered irreversible after age 3 [10]. Growth stunting and chronic wasting can be attributed to environmental enteric dysfunction (EED), which may be induced by chronic exposure to unsafe water [11].

Access to improved water. The 2010 DHS indicates that 89.6% of households have access to improved sources of water (piped water, public tap/standpipe, borehole, protected dug well, protected spring, rainwater or bottled water) while another 7% have access to a

unimproved source (unprotected dug well, unprotected spring, tanker truck/cart with drum or surface water) [12]. While this appears promising, research from Phase 1 of the intervention indicated that among the 468 households our team assessed for total coliforms across 19 districts in Rwanda, 27.8% of "improved" water sources, 80.2% of unimproved water sources and 58.3% of stored water supplies exceeded the Rwanda Standard for Potable Water [13]. 34.7% of households do not treat their water; among households that do, boiling is the dominant method (58.5%) [12].

Access to improved sanitation. In rural areas, approximately 25% of households countrywide use unimproved sanitation facilities, the majority of which are pit latrines without a slap/open pit. Another 57% in rural areas have access to an improved/not shared pit latrine with a slab [12]. As these figures were calculated country-wide, we might expect the proportion of households that only have access to unimproved facilities to increase in the lower wealth quintiles, although these data were not disaggregated by socioeconomic status. In addition, only 10% of households have a place dedicated for handwashing, and among those households, only 21% actually have soap and water for hand washing. This figure decreases in Western Province, where only 4% of households have a designated handwashing place [12]. 77% of households in the lowest wealth quintile have no water, soap or cleansing agent available [12].

Background of methods to assess exposures and disease

Water quality. Microbiological quality of drinking water can be assessed by enumerating total coliforms, thermotolerant coliforms (TTC) and *E. coli* cells in a water sample as an

indicator of overall contamination. “Total coliforms” refer to gram-negative, rod-shaped bacteria and include thermotolerant coliforms and bacteria of fecal origin. As not all coliforms are of fecal origin, though, the WHO has designated TTC (coliforms that specifically grow at 44 or 44.5°C) along with *E. coli* as a fecal indicator [14].

Microbial water quality can vary rapidly and widely, and peaks in pathogen concentration can cause outbreaks of waterborne disease. This is particularly relevant for households that collect their drinking water from surface water, springs or streams, since microorganisms can adhere to sediment and released *en masse* as water flow increases [15]. As water flow can be related to seasonal impacts of rain, our study incorporates the analysis of drinking water samples from all study households on a quarterly basis throughout the one-year follow-up period.

IMCI and Classifications of Diarrhea with Dehydration. Dehydration resulting from diarrheal disease is the principal contributor to the disproportionately high mortality rate attributed to diarrheal disease in Sub-Saharan Africa. In our study, enumerators will detect sunken eyes, inability to drink or limited drinking, eager drinking and restlessness and irritability. The skin pinch test will be applied and timed in order to classify dehydration. Lethargy and unconsciousness will also be assessed, as this is a danger sign associated with severe dehydration [16].

Assessment of malnutrition: Severe malnutrition disproportionately affects children in Sub-Saharan Africa and is linked to high child mortality. Weight-for-height *z*-(WHZ)

scores and mid-upper arm circumference (MUAC) have historically been used to assess childhood malnutrition and wasting in field settings [17]. Both MUAC and WHZ scores are comparable in their ability to predict 90-day mortality [17], [18]; however, MUAC is an inexpensive and straightforward method to assess nutritional status in field settings. For this reason, MUAC will be used to objectively assess childhood malnutrition in this study.

Justification of nested seroconversion study within the RCT

Seroconversion studies can elucidate the impact of improved water sources on personal contact with enteric pathogens. In addition, consistent contact with enteric pathogens and prolonged or recurrent diarrheal disease can subsequently result in intestinal enteropathy. This study will examine the impact of the intervention water filter on seroconversion against common enteric pathogens in pre-weaning children. These children will be followed up after a minimum of 6 months, at which point alternative food and water introduction has likely occurred.

Seroconversion study. Serological assays that assess antibody production against various enteric pathogens can provide a far more objective measure of exposure to enteric infections than reported diarrhea or diarrhea diagnosed using clinical indices in the field [19]. Multiplex immunoassay technology has recently been advocated as a convenient platform for surveillance of neglected tropical diseases and enteric infections [20]. Steinberg et al. (2004) sought to assess the age-specific seroprevalence of antibodies against various enteric pathogens, such as *E. coli*, Norovirus, *Cryptosporidium parvum* and *Helicobacter pylori* and hepatitis A virus (HAV) in order to determine an appropriate age

range to assess seroconversion against water and sanitation improvements in this population. Previous studies had demonstrated a marked age-specific prevalence of antibodies against these pathogens between 6 to 36 months of age [21–25]. As Table 2.1 depicts, Steinberg et al. found that for antibodies against *E. coli* heat-labile enterotoxin (ETEC-LT) and Norovirus, seroprevalence was lowest among children 6 to <12 months old compared with the 4 older age groups examined (12 to <18 months; 18 to <24 months, 24 to <30 months and 30 to <36 months). They observed the steepest increase in antibody acquisition between 6 and 18 months of age [26]. Antibodies against *C. parvum* increased with age, albeit more gradually, with seroprevalence peaking and leveling out at around 18 to 24 months of age. Prevalence of antibodies against HAV increased in the three oldest age groups, while antibodies against *H. pylori* remained fairly constant across all ages [26].

In a separate randomized controlled trial of household water treatment in Guatemala by Crump et al. (2007), households with children <12 months of age or a pregnant women in her last trimester of pregnancy were enrolled [19]. Households were randomly assigned to five different treatment groups: 1) flocculant-disinfectant; 2) flocculant-disinfectant plus a customized vessel; 3) sodium hypochlorite; 4) sodium hypochlorite plus a customized vessel; and 5) control (no treatment). Nested within this RCT was a serologic study that assessed age-specific prevalence of antibodies against *C. parvum*, ETEC, norovirus and *Giardia intestinalis* at baseline in children 6 to 12 months old and again at follow-up in the same children when they were 13-18 months of age. Neither the flocculant-disinfectant nor hypochlorite-disinfectant groups were associated with altered seroconversion after this 6-month period when compared to children in control households; however, serologic

response was associated with diarrhea prevalence. The authors surmise that seroconversion studies can still be useful in assessing health effects for household-based water treatment, particularly given this method's objective nature.

Moss *et al.* (2014) reported IgG responses in a cohort of Haitian children against various enteric protozoan pathogens: *Entamoeba histolytica* lectin adhesion antigen (LecA), recombinant *C. parvum* antigens 17- and 27-kDa (Cp17 and Cp27) and *G. intestinalis* variant-specific surface proteins (VSP1-VSP5) [27]. These novel targets were developed and validated for use in a prior study [28]. They also examined IgG responses against *Schistosoma japonicum* glutathione-*S*-transferase (GST), which is a control antigen included to account for non-specific binding. They also examined the temporal relationship between the detection of *G. intestinalis* and *E. histolytica* cysts in stool and the IgG response and found a sustained antibody response after 5 to 6 months. They found that IgG responses to all antigens studied were relatively low at 1 year of age, but that after that, IgG responses began to increase, with IgG responses peaking around 2 years of age before dropping off [27]. They also observed significant difference in antibody responses to LecA and VSP1-VSP5 antigens between rainy and dry seasons, with a similar (albeit non-significant) trend observed for Cp17 and Cp27 [27].

Sample medium for serological markers

Dried blood spots (DBS) are frequently used as a means for collecting blood for antibody sero-surveys and biomarker measurements in resource-limited settings. DBS samples can

be stable at room temperature for up to a week, alleviating concerns regarding storage conditions of biological samples in the field.

TropBio™ filter discs are to be used for the seroconversion study in conjunction with single-use retractable heel stick or finger lancets. TropBio™ filter discs have been used in previous field seroepidemiologic studies and are suitable for use with small children, from whom we require only a small amount of blood (50-60uL) [29].

Due to its high elevation (Figure 1.1), our study area exhibits a fairly temperate climate, with temperatures rarely exceeding 27°C. The average daily high temperature in Kibuye, Rwanda, which marks a central point of our study area, is 25°C, with nighttime temperatures reaching 14°C. While our samples will be protected from humidity and ultraviolet exposure, DBS samples are typically stable at the ambient outdoor temperatures seen in our study area. At the end of the day, all filter discs will be left on a counter to dry overnight. In the morning, they will be placed in plastic resealable bags with desiccant, with up to four samples in each bag, and stored in a closed box to protect the samples from UV exposure. All samples will be placed in a -20°C freezer at the NRL within 8 days of sample collection.

We intend to analyze the blood samples for the seroconversion study on the Luminex xMAP platform at the U.S. Centers for Disease Control and Prevention (CDC); as such, they will have to be exported to the United States under a Material Transfer Agreement (MTA) with the Ministry of Health. In this case, DBS samples are considered exempt from

International Air Transport Association (IATA) shipping regulations, as they are considered non-infectious once dried.

Age Group in months	ETEC	Norovirus	*C. parvum*	HAV	*H. pylori*
6 to <12	48%	27%	27%	40%	20%
12 to <18	81%	61%	53%	28%	19%
18 to <24	80%	83%	70%	46%	21%
24 to <30	77%	94%	67%	60%	25%
30 to <36	83%	94%	73%	76%	25%

Table 2.1. Seroprevalence of antibodies against various enteric pathogens, by age group in months, in San Juan Sacatepequez, Guatemala. (Steinberg et al., 2004 [26])

Chapter 3 – Pneumonia and Household Air Pollution

Acute lower respiratory infection (respiratory disease) is among the leading causes of morbidity and mortality among children under-5 years-old in Sub-Saharan Africa and is responsible for 4.6% of global disease burden [1]. Respiratory disease is the 3rd largest contributor to years of lives lost in eastern Sub-Saharan Africa, behind only HIV/AIDS and malaria [1]. Household air pollution (HAP) alone is associated with 3.5 million deaths annually and 4.3% of all global DALYs in 2010, making it the 3rd leading risk factor for global burden of disease [2]. HAP is the primary contributor of respiratory disease in children under 5 and a leading cause of chronic bronchitis and chronic obstructive pulmonary disease (COPD) among women [2,3]. Among the top non-communicable contributors to global DALYs, HAP is the only exposure that disproportionately affects children [2]; HAP is directly associated with ALRI in addition to low birth weight and an increased risk of tuberculosis [30]. While WHO guidelines recommend that household levels of HAP based on particulate concentrations not exceed 10 ug/m^3 for $PM_{2.5}$ and 50ug/m^3 for PM_{10}, households that burn biomass for cooking, concentrations can exceed 10,000 ug/m^3 during cooking or other periods [3]. In addition to the direct link between HAP and development of lower respiratory infections, HAP has been associated with inflammatory processes linked to cardiovascular sequelae.

Many of the poorest households in our study area cook using biomass fuels on inefficient cookstoves. Given the high burden of disease attributable to these exposures in our study population, minimizing exposure to biomass fuels could have wide-ranging effects on health and survival [31]. Despite demonstrated HAP reductions in previous cookstove trials

[32], however, few trials have demonstrated a clear link between improved cookstove distribution and pneumonia reduction.

Cooking habits and fuel sources. According to the 2010 DHS, 77% of households use wood as cooking fuel country-wide, with the 2nd most common cooking fuel being straw, shrubs or grass (12%); however, the proportion of households that use wood for cooking fuel increases to 83.3% in rural areas, with 12.4% using straw, shrubs or grass and 3.0% using charcoal [12]. Households that use wood fuel for cooking spend, on average, 50 minutes per day collecting wood. Most in Western Province report that they prefer wood to other sources because it is relatively available (72%) and because it can be obtained without purchasing (17%) [33]. The predominant difficulties reported by residents in Western Province with gathering wood are that the activity is difficult in the rainy season (24%) and that cutting wood is difficult (21%). Other reported difficulties included that gatherers had to travel too far to get firewood, that the wood is heavy and that the wood is sometimes located in inaccessible areas [33]. In rural areas, 28.9% of cooks report cooking inside the house, 52.2% report cooking in a separate building, and 18.0% report cooking outside [12].

Disease outcome misclassification and recall bias could be attributed to subjective nature of self-reporting inherent in household questionnaires. This is a particular problem in assessing the impact environmental heath interventions, most of which cannot be blinded. Field-based epidemiologic studies, clinical trials and monitoring and evaluation of interventions to prevent respiratory and enteric infections—two of the major killers of young children in Rwanda and worldwide—rely largely on subjective assessments of

reported symptoms. For example, the WHO's IMCI criteria for pneumonia case identification in resource-limited settings entails disease indices that incorporate cough, difficulty breathing and rapid respiration [34]. As a result of subjective methods of case identification, previous clean cookstove trials and studies of household air pollution (HAP) and acute lower respiratory infection (ALRI) prevalence have yielded considerable statistical heterogeneity which is ascribed in part to methodological differences in exposure and disease ascertainment [31,32,35].

By coupling exposures and disease outcomes with objective, quantitative biomarkers, the physiologic effects of cookstove interventions can be better measured and described. Other health outcomes, such as ALRI or chronic lung conditions and cardiovascular sequelae, such as hypertension, in primary household cooks will also be explored in this study, although biomarker data may provide a crucial objective link between environmental exposures and disease.

Background for household air pollution exposure assessment methods

Under perfect combustion, only CO_2 and water would be produced from burning wood fuel; however, traditional cookstoves, such as the three-stone fire commonly used in Rwanda, has a combustion efficiency of only 20-30%. As a result, indoor smoke can contain potentially harmful pollutants, such as particulate matter (e.g., PM_{10}, $PM_{2.5}$, fine PM), nitrous oxide (N_2O), sulfur oxides, carbon monoxide (CO), formaldehyde and various carcinogens, such as benzo[a]pyrene, benzene and various other polycyclic aromatic hydrocarbons (PAHs) [36]. Because of the harmful chemicals and particles emitted by

wood smoke, reducing exposure to HAP should be the principal objective of any household energy intervention [36]. Objective HAP measurements should be performed in order to accurately measure emissions and determine whether or not the intervention has driven reductions in emissions. Several study design considerations have to be applied for HAP exposure assessment, such as duration of monitoring and seasonality [37]. Previous trials have monitored HAP from 24 hours up to 7 days. In general, it is better to monitor HAP over longer periods of time, given that there may be some variation in day-to-day behavior [37]. Seasonality can affect exposure, as the intervention specifically instructs recipients to use their cookstove outdoors (or in a doorway during rain); however, during the rainy season, cooks may move their stove indoors, thereby increasing exposure.

In general, the most accurate way to determine personal HAP exposure is personal HAP monitoring over a 24- or 48-hour period [38]. Particularly vulnerable groups, such as women and children, can be more thoroughly assessed using this method. Carboxyhemoglobin (COHb), as a biomarker, has been used in previous cookstove studies and can reflect recent exposure to cookstove smoke within the previous few hours [38]. Due to the increased accuracy of personal household exposure monitoring, we will be using 48-hour assessments of personal exposure using a pump and filter system programmed so that 24-hour mean $PM_{2.5}$ concentration can be calculated. In addition, we will assess COHb levels through pulse oximetry and exhaled CO.

As Figure 3.1 depicts, some regard biomarkers as the most accurate method to determine personal exposure. While it is considered the most expensive method, costs can potentially be mitigated by using more cost-effective means of collecting and storing samples.

Field diagnosis of recent and current pneumonia cases

The primary caregiver of each child under-5 will be asked about symptoms of respiratory disease in the previous week, including constant cough, difficulty breathing/panting/wheezing, rapid breathing and the presence of a blocked or runny nose. In addition, IMCI criteria will be used to identify any potential child health danger signs and to differentiate between upper respiratory infections and pneumonia. While IMCI criteria can be useful to diagnose pneumonia in resource-limited settings, they can still be somewhat subjective. Chest indrawing and stridor can be difficult to detect in some children, and breath counts (which must be taken over the course of a minute) can be difficult to obtain in very young or restless children. For these reasons, objective and quantitative methods should be developed that are easily applied in the field to mitigate the risk of disease misclassification. These methods will be subsequently applied soon after the dissertation is submitted; however, because of the immediate relevance, the background and justification for these methods is provided in the next section.

Biological mechanisms of HAP-associated pneumonia pathogenesis and justification of biomarkers for future cookstove studies

Recognizing the potential contribution of biomarker data, there are increasing calls for such data to be incorporated into future clean cookstove trials [7,8]. While funding limitations

prevented biomarkers from immediately being incorporated into this dissertation, blood samples have been collected during baseline and round 2 of follow-up with the intention of analyzing these samples in the near future. We anticipate assessing levels of various inflammatory biomarkers (including interleukin (IL)-6, IL-8, IL-10, tumor necrosis factor-alpha (TNF-α) and C-reactive protein (CRP) for the cookstove component. We also anticipate analyzing carboxyhemoglobin (COHb) as an indirect measurement of carbon monoxide (CO) exposure, which is a key contaminant emitted during wood combustion. Inflammatory biomarkers can be indicative of several disease processes associated with $PM_{2.5}$ exposure, particularly ALRI and hypertension. Both outcomes are being individually assessed in our study. Meanwhile, COHb levels exceeding 2.5% have been deemed unsafe by the WHO, and excess COHb levels (>5%) are associated with neurobehavioral factors, impaired vision and decreased alertness, in addition to potentially being an indicator of overall HAP exposure [30].

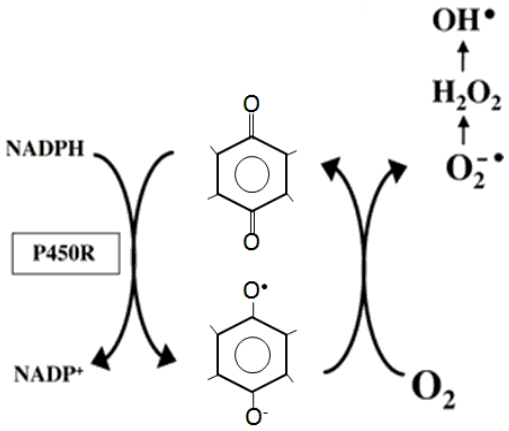
$PM_{2.5}$ induces reactive oxygen species (ROS) generation due to the presence of redox-active transition metals, redox cycling quinoids and polycyclic aromatic hydrocarbons that may be present on the surface of these particles. Iron is particularly redox-active, with a high propensity to flip between different valence states [39], and is frequently present on the surface of these particles [40]. Iron-induced ROS production is catalyzed through Fenton-type reactions [39]:

$$Fe^{3+} + O_2 \bullet \rightarrow \mathrm{Fe}^{2+} + O_2$$

$$\underline{Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH + \bullet \mathrm{OH}}$$

$$O_2 \bullet^{-} + \mathrm{H_2}O_2 \rightarrow_{Fe} OH^{-} + \bullet \mathrm{OH} + \mathrm{O_2}$$

PAHs and various redox-cycling quinones are also present on the surface of PM, and they are highly persistent [40]. In the presence of nicotinamide adenine dinucleotide phosphate (NADPH) or ascorbate, P450 reductase catalyzes the one-electron reduction of a carbonyl oxygen of a quinone, converting it into a highly reactive semiquinone radical. This radical can then react with O_2, which can induce the release of $O_2^{-}\bullet$ radicals [41].



Fine PM is capable of penetrating deep into the lung and into alveoli, which can induce an endogenous inflammatory response which can cyclically induce more ROS production. Given that this ROS production is occurring in the gas exchange regions of the lung, ROS and inflammatory cytokines can dissolve into the bloodstream [40]. This can have a fairly acute impact on the blood vessels in addition to a more chronic effect on cardiovascular and respiratory health.

Biomarkers of indoor air pollution. Relatively few studies have focused on linking cookstove exposure to levels of systemic inflammation in addition to biomarkers of exposure. A study by Torres-Dosal et al. (2006) assessed levels of COHb one month after a HAP intervention to biomass-using households [30]. The intervention consisted of indoor soot removal, paved dirt floors and a sand, clay and cement Patzari stove with a chimney.

They found no difference in COHb between intervention and control groups one month after intervention deployment [30]; however, their sample size was fairly small (n=20) and massive COHb reductions were observed across both study arms, indicating that some other exposure or potential confounding variables were not being properly accounted for.

A couple of studies in India demonstrated reductions in systemic inflammation in women who received liquid petroleum gas (LPG) stoves to reduce HAP. In one study, significant reductions were observed for median and mean TNF-α, IL-6, IL-8 and IL-12 levels among women cooking with LPG stoves compared to women who cooked exclusively with biomass [42].

Dutta et al. (2012) examined the relationship between HAP and levels of CRP, IL-6, IL-8 and TNF-α in the context of also assessing potential linkages between HAP, pro-inflammatory markers and blood pressure [43]. They found a positive association between all four pro-inflammatory mediators, as measured by ELISA, and household levels of PM_{10} and $PM_{2.5}$, as measured in the cooking area by a real-time laser photometer, after controlling for age, BMI, education, family income and kitchen location. They also compared levels of these pro-inflammatory markers between women who cook with biomass vs. LPG stoves. Compared to women cooking with LPG stoves, women cooking with biomass also had significantly higher levels of TNF-α (29.4 +/- 8.1 vs. 18.1 +/-5.4 pg/mL, $p<0.0001$), IL-8 (29.6 +/- 5.4 vs. 12.2 +/-5.6 pg/mL, $p<0.0001$), IL-6 (5.6 +/- 2.2 vs. 1.6 +/-0.5 pg/mL, $p<0.0001$) and CRP (6.2 +/- 1.3 vs. 1.9 +/- 0.7 mg/L) [43].

While this study reveals an interesting association between HAP, stove type and levels of pro-inflammatory markers, it was cross-sectional and did not account for how long LPG users had been using their stove. Importantly, some demographic factors, such as years of schooling and family income, were not balanced between LPG and biomass groups. Finally, while LPG stoves are highly efficient and do not involve biomass, LPG can be expensive and logistically unfeasible in many resource-constrained settings. It is for this reason that more efficient biomass-burning rocket stoves are being investigated in this study, as these stoves may be more feasible to bring to scale. The use of biomarkers of systemic inflammation has not been thoroughly investigated with these types of stoves.

Inflammation and ALRI. Given that inflammation in the deep lung can induce increased alveolar permeability, exposure to air pollutants can exacerbate ALRI and worsen pneumonia prognoses. Complex interactions between cytokines and chemokines directly mediate, amplify and maintain lung injury processes [44]. TNF-α and IL-1β are crucial early response cytokines in lung inflammation, and both cytokines are typically present in the bronchoalveolar lavage fluid (BALF) of patients at risk for acute respiratory distress syndrome [44,45]. Levels of TNF-α, IL-8 and IL-6 all appear to be elevated in the BALF of patients with severe pneumonia and acute respiratory failure compared to healthy controls [45]. Sack et al. (2005) collected exhaled breath condensate (EBC) from intensive care unit patients with severe pneumonia and acute lung injury (ALI) and from smoking and non-smoking health volunteers [46]. The EBC was analyzed for levels of several cytokines via a multiplex cytometric bead assay (CBA). They found that the cytokines IL-

1β, IL-6, IL-8, IL-10, IL-12 and TNF-α were all significantly elevated in the ICU patients with ALI compared to health smokers and health non-smoking volunteers ($p<0.01$) [46].

Among an inpatient cohort of pneumonia patients, mean IL-6 concentrations were higher among patients who developed sepsis vs. those who did not ($p<0.001$). IL-6 levels also appeared to be an indicator of sepsis-associated mortality, as those who died from pneumonia-induced sepsis had higher levels of IL-6 than those who survived ($p<0.003$) [47]. While TNF and IL-10 concentrations were lower than IL-6 concentrations overall, pneumonia patients with severe sepsis still had significantly higher levels of TNF and IL-10 [47]. In another study, TNF-α and IL-10 was only observed to be elevated among patients with acute respiratory distress syndrome (ARDS) and were not detected among normal healthy controls [44]. Patients with diagnosed ARDS had IL-6 levels that were 100-fold higher than in normal subjects ($p<0.005$). The authors also found that day 7 concentrations of IL-1β and IL-6 and the ratio of IL-6 to soluble IL-6 receptor (sIL-6R) in ARDS patients were significantly associated with subsequent mortality ($p<0.05$) [44].

Taken together, serum cytokines may be indicative of the presence of pneumonia, particularly if elevated cytokines are present with other ALRI symptoms. Cytokine levels can provide crucial information in resource-limited settings, in which access to health care may be limited and advanced on-site diagnostic approaches are unavailable. In addition, cytokines may reflect overall disease prognosis and may indicate mortality risk.

Inflammation and hypertension. Inflammation affects cardiovascular health primarily through mediating vasoconstriction and atherosclerotic processes associated with hypertension. ROS, which can induce inflammatory mechanisms described previously, may directly affect the integrity and tone of the vascular endothelium. Endothelium cell (EC)-derived vasodilators, such as nitric oxide (NO), prostacyclin (PGI2), and vasoconstrictors, such as ET-1, thromboxane and angiotensin II (AngII) are typically balanced to prevent excess vasodilation or vasoconstriction from occurring. Inflammation can disturb this balance of EC function, more often favoring the production of EC-derived vasoconstrictors [48]. Increased production of AngII and cytokines such as CRP, which may be induced by ROS, may in turn stimulate the production of more ROS. Excess ROS can stimulate NADPH oxidases and NO synthase which can result in mitochondrial dysfunction, further stimulating ROS production. This can ultimately induce pro-inflammatory transcription factors nuclear factor-kappa B (NF-κB), Nrf2 and AP1 [49]. These transcription factors can then bind to DNA and upregulate target pro-inflammatory cytokines such as IL-6, monocyte chemoattractant protein (MCP-1) and TNF-α [49,50]. Neoantigen formation in the blood induces T-cells to release IL-17, which promotes the entry of macrophages and other inflammatory cells into the bloodstream. While the mechanism is ill-defined, pro-inflammatory cells in the bloodstream can induce vasoconstriction and water and sodium retention, which can contribute to hypertension if this process is not controlled [49].

Cytokines can also induce vascular cell growth and migration. IL-6 and TNF-α upregulate vascular endothelial growth factor (VEGF), while IL-1 can trigger abnormal proliferation

of vascular smooth muscle cells (VSMCs) [48]. This outlines another mechanism by which ROS can induce vascular stiffening and vasoconstriction. While ROS-induced hypertension may be linked to multiple pathways, all pathways are linked to elevated CRP levels. In theory, if CRP-induced inflammatory responses can be controlled, inflammation-mediated hypertension can be reduced.

Dutta et al. (2012) were principally interested in determining whether or not the pro-inflammatory cytokines CRP, TNF-α, IL-6 and IL-8 were associated with both HAP exposures and cardiovascular disease [43]. After assessing the relationship between cookstove type and levels of these markers, 91.2% of women cooking with biomass had serum CRP levels exceeding 3.0 mg/L, designated as the cut-off value for cardiovascular risk; meanwhile, among women cooking with LPG stoves, only 12.4% were considered high risk. In addition, systolic blood pressure (SBP) and diastolic blood pressure (DBP) were positively correlated with IL-6 and SBP was positively associated with CRP after controlling for potential confounders. Meanwhile, a positive correlation was observed between CRP levels and tachycardia [43].

This study will incorporate objective measures of exposure, such as personal HAP exposure measurements, carboxyhemoglobin concentrations and arterial oxygen saturation, to assess the effect of stove types and cooking exposure on $PM_{2.5}$ and the relationship between $PM_{2.5}$ and carbon monoxide exposures. Outcome measures, such as self-reported child upper respiratory infection and IMCI-identified ALRI cases, will be

assessed relative to various stove types and cooking behaviors. The biomarker component of this study is of interest and will be incorporated at a later date.

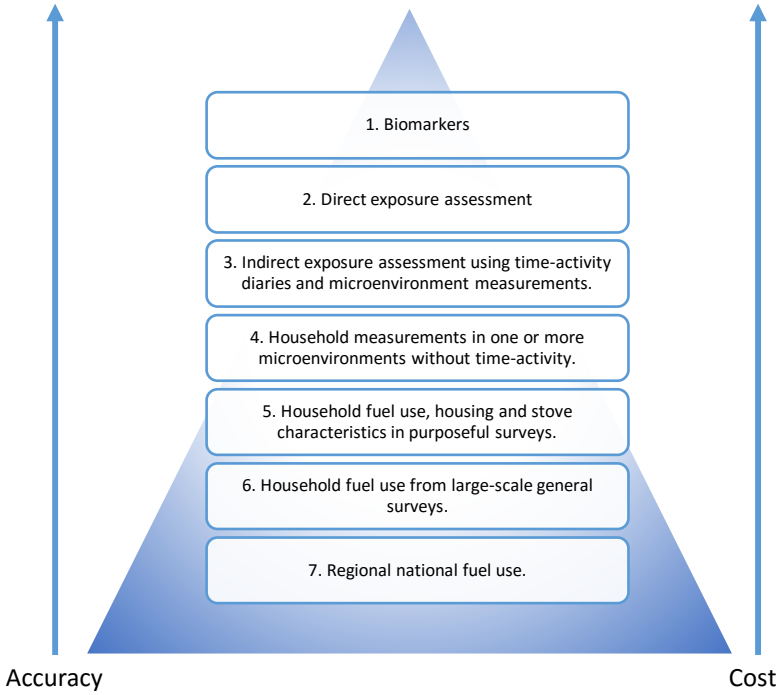


Fig. 3.1. Continuum of accuracy and cost of different methods of household air pollution assessment, adapted from the WHO Workshop for Indoor Air Pollution and Household Energy Monitoring. [36]

Chapter 4 – Study Context, Preliminary Data and Unifying Methods for Baseline Cross-Sectional Study

Study Site, Population and Household Characteristics

Western Province is located in western Rwanda, bordered by the Democratic Republic of the Congo and Lake Kivu to the west, Burundi to the south, Southern Province (Rwanda) to the east/southeast and Northern Province (Rwanda) to the east/northeast (Figure 1). The province is divided into 7 districts: Karongi, Rutsiro, Rubavu, Nyabihu, Ngororero, Rusizi and Nyamasheke. These 7 districts are further divided into 96 administrative sectors, which in turn are divided into 538 cells and 3612 villages. Our study population will focus on households classified as the poorest 30% of the population, designated by the Rwandan government as *Ubudehe* levels 1 and 2, with children under-5 years old.

Geographic distribution of population. According to the 2012 Population Housing Census, approximately 10.5 million people live in Rwanda, reflecting 25% population growth since 2002. Of this total population, 2.5 million reside in our study area, Western Province. The population in each district ranges from about 296,000 in Nyabihu district to over 404,000 in Rusizi and Rubavu districts [51]. While Rusizi and Rubavu are the most populous, their populations are disproportionately distributed to urban areas (Cyangugu in Rusizi District and Gisenyi in Rubavu District) compared to other districts. Gisenyi and its surrounding areas, in particular, were not drawn into our overall study population since households in these areas use coal, rather than wood biomass, for cooking (Figure 1.1b). Because their

fuel source differs from those residing in more rural areas, they were not eligible to receive the intervention stove, which is intended to use wood.

Education. Western Province has the highest percentage of men (17%) and women (26%) who have never obtained any formal education; however, this figure is largely influenced by age. Across the country, for example, 79% of women age 65 and over have no education, compared to 2% of girls between the ages of 10 and 14. We anticipate the same trend in our study population: the older our respondents are, the less likely they are to have had a formal education.

Intervention information. During Phase 2 of DelAgua's *Tubeho Neza* program, the paired intervention was distributed at distribution events held at cell offices, which typically administer to 4-5 villages. The point-of-use water filter distributed was the Vestergaard Frandsen LifeStraw Family 2.1 (Figure 4.1). This filter is a point-of-use microbial water treatment system, ideal for household use in resource-limited settings. The intervention consists of a 20 micron filter, through which the water is initially poured. This water is then filtered again through a 0.20 micron hollow-fiber ultrafiltration membrane into a 5.5 liter safe storage container. The filter is equipped with a plastic tap, which eliminates the need to open the storage container. This system filters up to 18,000 liters of water throughout its lifetime, which can supply a family of five with clean drinking water for three to five years. This system exceeds the World Health Organization's "highly protective" standard for household water treatment technologies [52].

The cookstove that will be distributed during this trial is the EcoZoom™ Dura, which concentrates the combustion process while channeling air flow to completely burn the combustible fuel provided (Figure 4.2). The carbon-rich fuel (typically wood) is placed in a ceramic combustion chamber within steel housing, which is topped with a cast iron stove. The stove includes a rack for placing small pieces of firewood. The cookstoves and water filtration units will be delivered together at no cost to participating households during village meetings.

Households receiving the cookstove and filter were introduced to their projected health benefits and filter and cookstove use and maintenance at the cell-level distribution events. A set of hired community health workers followed up with each receiving household within 48 hours of the distribution to distribute pictorial brochures and posters and to properly train all household members on the proper use and maintenance of the filter and stove.

Preliminary Data from Pilot Phase and Phase I

In order to lay the foundation for a large-scale intervention, DelAgua Health and the Ministry of Health undertook a pilot project in Western Province starting in 2012. Designated as “Phase 1”, the intervention consisted of the distribution and promotion of Lifestraw™ water filters and EcoZoom™ improved-efficiency cook stoves to 1943 households across 15 villages in 11 districts [5] in Western and Southern provinces. The objective of this phase was to determine whether certain design criteria, in conjunction with the cookstove and water filter intervention, could result in increased uptake and use of the stoves and filters. The program design choices that DelAgua selected were 1) Free distribution; 2) Behavior change communication through community- and household-

based activities; and 3) “Pay for Performance” public-private partnership between DelAgua and the Rwandan Ministry of Health, with performance incentivized by carbon credit revenues.

The methods of the pilot phase of the study are described in detail by Barstow et al. (2014) [5]. This phase utilized a convenience sample of 15 non-randomly selected villages in 11 districts. It applied behavior change theory, particularly the Diffusion of Innovation theory [53] and the Health Belief Model [54], to encourage uptake of the intervention and engagement with community members, as described previously. Ministry of Health Environmental Health Officers (EHOs) conducted surveys in 325 households each month to assess demographics, cooking practices and water treatment and collection practices. Qualitative interviews were also performed at focus group sessions jointly conducted by CHWs and EHOs which addressed the perceived adoption of the intervention within their villages, problems with the filter and stove, effectiveness of household messaging and barriers to exclusive use of the filter and stove. The results of the pilot have been published [5,6,55] and describe uptake, use and the environmental impacts of the stove and filter in the home.

Uptake and Use. Reported and observed measures indicated that the filters and stoves were largely embraced by most of the target population in the pilot phase. Adoption of the water filter met or exceeded 90% on several metrics. 96.5% of households reported use, while visual inspection of the water filters indicated that 9 out of 10 household water filters were filled with water (the metric EHOs used to visually confirm filter use). 12.8% of

households also reported using the water filter for reasons other than providing safe drinking water, such as cooking, hand washing and washing dishes [5]. Reported primary use of the EcoZoom stove was around 90% as well; however, 71.5% of households reported concurrent use of their traditional stove. About 21% of households were cooking at the time of the follow-up visit, and of those, 63.7% were using the EcoZoom stove, 21.9% were using a traditional 3-stone fire, 11.4% were cooking using a different type of stove, and 2.9% of households were cooking concurrently with the EcoZoom stove and a traditional stove. As households were also counseled to cook outside, cooking locations were also assessed. At the follow-up visit, 37.3% of *Ubudehe* 1 and 2 households surveyed were cooking outdoors, 44.5% were cooking indoors, and 18.2% were cooking in a separate kitchen detached from the main living structure.

Methods for assessing impact of intervention on household air and water quality. The evaluation of the pilot occurred from September 2012 to April 2013. The purpose of the evaluation was to assess uptake and use of the filter and stove and their impact on environmental exposures (drinking water quality and HAP); it also sought to provide information that could assist in the design of a comprehensive health impact evaluation that was envisioned for Phase 2. $PM_{2.5}$ and TTC reductions were also assessed. The study was powered on reductions of $PM_{2.5}$ emissions rather than TTC counts since the former required a larger sample size. Ultimately, 63 households from three rural villages were required for each study arm to assess $PM_{2.5}$ and TTC reductions. These households were drawn from households included in the larger pilot study.

During each of the five monthly visits, field enumerators collected water samples from the primary drinking water container in sterile Whirl-Pak bags (Nasco, Fort Atkinson, WI) for microbiological assessment using membrane filtration. All water samples were placed on ice in the field and processed within 6 hours of collection. The membrane filtration technique was performed on membrane lauryl sulphate medium (Oxoid Limited, Basingstoke, Hampshire, UK) and samples were incubated using a DelAgua field incubator (Robens Institute, University of Surrey, Guilford, Surrey, UK).

Unifying Methods for Dissertation Aims

Principal Components Analysis to derive socioeconomic index for Aims 1 and 3 (Chapters 5 and 7). Our team collected data on a total of 17 socioeconomic and demographic variables that were not direct exposures of interest. In order to create a summary socioeconomic indicator, we performed a principal component analysis (PCA) to create an overall socioeconomic index. The application of PCA methods to discrete socioeconomic indicators, as suggested by Filmer & Pritchett (2001), is problematic as it assumes a normal distribution of all underlying variables. In addition, dichotomizing categorical variables into binary dummy variables leads to loss of ordering inherent in all ordinal variables. Generally, the Filmer-Pritchett method for PCA leads to an underestimation of the reported proportion of explained variance as the number of categories increase, and greater household misclassification generated SES index ultimately calculated from the factor scores [57]. In order to apply PCA to our socioeconomic and demographic variables to attain an SES index, we used original ordinal variables rather than binary dummy variables

and input a polychoric correlation matrix between the variables into the PCA to account for the ordinal and binary nature of asset, education and housing construction variables.

Variables were not considered for PCA if univariate analysis revealed a standard deviation (SD≤1) and if the population was fairly homogenous across that variable (≥95%). As a result, the following variables were removed from consideration of inclusion into the PCA: "Ownership of television," "ownership of bicycle," "ownership of motorcycle," and "ownership of boat." High collinearity was present between the following three variables: "Household owns land/plot not used for agriculture," "household owns agricultural land." In this case, only the variable for "household owns agricultural land" was retained for the PCA since it had the highest standard error and thus would contribute more variability than the other collinear variable.

The following variables were included in our analysis: Household head education (0=None; 1=Nursery; 2=Some primary; 3=Completed primary; 4=Some secondary; 5=Completed secondary), primary cook education (0=None; 1=Nursery; 2=Some primary; 3=Completed primary; 4=Some secondary; 5=Completed secondary), household has electricity (1=yes, 0=no), ownership of radio (1=yes, 0=no), ownership of mobile telephone (1=yes, 0=no), ownership of mattress (1=yes, 0=no), ownership of agricultural land (1=yes, 0=no), ownership of house (1=yes, 0=no), cares/tends cows (1=yes, 0=no), type of flooring materials (0= earth/sand or animal dung, 1=bricks, 2=cement, type of wall materials (0= only wood planks, 1=wood planks and mud, 2=mud bricks- not covered, 3=mud bricks-covered with mud, 4=mud bricks-covered with cement, 5=real bricks-not

covered, 6=real bricks-covered with cement, 7=wood-covered) and roof materials (, 0=plastic sheets, 1=metal sheets, 2=clay tiles).

Principal Components Analysis to derive socioeconomic index for Aim 2 (Chapter 6)

Water source and sanitation type variables are frequently included in principal components analyses [58], but since these variables were primary exposures of interest, they were not incorporated into the PCA for Aims 1 and 3. In addition, the PCA for Aims 1 and 3 incorporated roof type, which is an exposure of interest in Aim 2 given its effect on ventilation. Therefore, the PCA was repeated for Aim 2 and included these water and sanitation variables and excluded roof type. A total of 16 socioeconomic and demographic variables were included that were not direct exposures of interest to either the baseline study for the water filter or cookstove RCT. Additionally, information was gathered on water supply and sanitation facility, which were primary exposures of interest for the baseline study of the water filters, but were consolidated with the other socioeconomic and demographic variables for this principal components analysis (PCA) to create an overall socioeconomic index. Like the first PCA, we used original ordinal variables rather than binary dummy variables and input a polychoric correlation matrix between the variables into the PCA to account for the ordinal and binary nature of asset, education, housing construction and water and sanitation variables, variables deemed as relevant to deriving socioeconomic indices in developing countries [58].

The following variables were included in our analysis: Household head education (0=None; 1=Nursery; 2=Some primary; 3=Completed primary; 4=Some secondary;

5=Completed secondary), primary cook education (0=None; 1=Nursery; 2=Some primary; 3=Completed primary; 4=Some secondary; 5=Completed secondary), household has electricity (1=yes, 0=no), ownership of radio (1=yes, 0=no), ownership of mobile telephone (1=yes, 0=no), ownership of mattress (1=yes, 0=no), ownership of agricultural land (1=yes, 0=no), ownership of house (1=yes, 0=no), cares/tends cows (1=yes, 0=no), type of flooring materials (0= earth/sand or animal dung, 1=bricks, 2=cement, type of wall materials (0= only wood planks, 1=wood planks and mud, 2=mud bricks- not covered, 3=mud bricks-covered with mud, 4=mud bricks-covered with cement, 5=real bricks-not covered, 6=real bricks-covered with cement, 7=wood-covered), water source (0=stream, river, pond, lake or rainwater, 1=unprotected dug well or unprotected spring, 2=protected dug well or protected spring, 3=public tap/standpipe or hand pump/borehole and 4=piped water into dwelling, yard or plot) and toilet type (0=No toilet/bush, 1=pit latrine with or without a slab, 2=ventilated pit latrine and 3=composting toilet or pour/flush toilet).

Accounting for non-normality of continuous variables. Univariate analyses were performed on all continuous exposure and outcome variables relevant to our analysis to assess normality, including age, TTC counts, $PM_{2.5}$ exposure (ug/m^3) and median fluorescence intensity values (MFI) for all serological responses to enteric pathogens. For water quality, TTC counts designated "too numerous to count" (TNTC) were right-censored and assigned the value of 300 CFU/100mL, deemed the maximal countable value [59]. All TTC counts were adjusted to a standard sample volume of 100mL. After these adjustments, univariate analysis of TTC counts revealed a U-shaped logit-normal distribution, with bounded outcomes at 0 and 300 CFU/100mL. Water quality was then stratified in accordance with

WHO-designated risk profiles: 0 CFU/100 mL (conforms to WHO guidelines; 1-10 CFU/100mL (low risk); 10-100 CFU/100mL (intermediate risk); and 100-1000 CFU/mL (high risk) [60]. In order to account for stratification and clustering of our survey data, analyses using generalized estimating equations will be performed on marginal models of this newly created ordinal response variable for water quality. $PM_{2.5}$ exposure was successfully log-transformed and incorporated into linear regression analyses. Serological indicators were transformed with varying degrees of success, and these procedures are described in detail in Chapter 7.

Sample weight construction. Since the intervention was randomized at a 3:1 ratio while intervention and control villages were selected at a 1:1 ratio, control villages were oversampled compared to intervention villages. Also, while an average of 9.2 households were selected per cluster, the number of households selected per cluster ranged from 2 to 11 households. To account for population proportional selection, stratification by treatment arm and different selection probabilities within each cluster, we calculated study weights for household-level and child-level observations. For household-level observations, we used the following formula, where p_{1si} is probability of selection of any given cluster *i* in treatment stratum *s* and p_{2sij} the probability of selection of household *j* within cluster *I* :

$$p_{1si} = c_{si}x / \sum_{i=1}^{n} c_s \quad ; p_{2sij} = n_{sij}/c_{si}$$

$$PS = p_{1si}\, X\, p_{2sij}; w_{sij} = \frac{1}{PS}$$

Where *x* is the number of clusters per stratum, c_{si} is the size of cluster *i* in stratum *s*, n_{sij} is the size of household *j* in cluster *i* in stratum *s*, PS is the probability of selection and w_{sij} is

the sample weight attributed to household *j* in cluster *i* within stratum *s*. Since all children under-5 were enrolled in each household, similar sample weights were calculated for child-level data, scaled by 1.45 , the anticipated average number of children under-5 present in every household that has at least one child under-5 in Western Province. This factor was calculated using 2012 Rwandan Census [61] and 2010 Demographic Health Survey [12] datasets.

Fig. 4.1. The LifeStraw Family 2.1 water filter, used as the intervention filter in this study.

Fig. 4.2. The EcoZoom Dura, used as the intervention cookstove in this study.

Chapter 5 - Predictors of diarrhea and coliform contamination of drinking water; a cross-sectional study in Western Province, Rwanda

Abstract

Diarrheal disease is one of the largest contributor to childhood morbidity and mortality in Sub-Saharan Africa. A large proportion of diarrhea attributable to inadequate access to water and sanitation occurs on the continent. We leveraged the baseline phase of a large cluster-randomized controlled trial of a point-of-use household water filter to characterize household and environmental exposures that are disproportionately associated with diarrheal disease among children under-5 years-old in Western Province, Rwanda. Taylor series linearization with robust variance estimation was used to account for the complex survey design of this study. While WHO JMP-designated improved water sources did not appear to be significantly protective over surface water sources and unimproved water sources, further disaggregation found that piped water and dug wells were significantly more protective than protected spring water sources, while standpipes, boreholes and surface water posed significant excess risk. WHO JMP-designated improved sanitation did not appear to be protective, but further disaggregation into actual toilet types revealed that composting toilets nearly doubled the risk of recent diarrhea. Both shared sanitation and toilet area cleanliness (measured as observed fecal matter within 1 meter of the toilet) increased seven-day diarrhea prevalence by 54% (RR: 1.54, 1.13-2.10) and 35% (RR: 1.02-1.80), respectively. While water quality, measured by thermotolerant coliform counts in drinking water samples taken from the household, was not associated with diarrhea, water quality results varied greatly by various household water and sanitation categories. Taken

together, these results reveal lessons for implementers and future priorities for water and sanitation researchers.

Introduction

Diarrheal disease accounts for approximately 3.6% of the total global disease burden [1]. Out of 1.3 million annual deaths attributable to diarrheal disease globally [62], approximately 502,000 and 280,000 deaths were associated with inadequate drinking water and sanitation, respectively [63]. While Sub-Saharan Africa only represents approximately 7.8% of the world's population, 45.7% of all water-attributable diarrheal disease mortality occurs on the continent. While global diarrhea attributable to inadequate water and sanitation has declined since 1990, it remains among the top contributors to childhood mortality on the continent [64]. The WHO Joint Monitoring Program's (JMP) review of global progress towards water and sanitation targets within the Millennium Development Goals (MDGs) reveals that while 43% of people in Sub-Saharan Africa have gained access to an "improved" drinking water source since 1990, over 50% of the population still lacks access [60]. Even the community-level improved drinking water sources considered to be "improved," recently acquired by 1.2 billion people globally, may be subject to detectable contamination by fecal pathogens [65], resulting in a potential over-estimation of protection conferred by improved drinking water sources in low- and middle-income countries (LMICs) as described in the 2010 Global Burden of Disease (GBD) study [2]. A systematic review of 345 water quality studies by Bain et al. (2014) revealed that in rural areas throughout Sub-Saharan Africa, 58% (95% CI: 0.41-0.71) of piped water sources, 22% (0.15-0.31) of boreholes and 91% (0.82-0.96) of unprotected groundwater sources were contaminated with fecal indicator bacteria (FIB), defined as any detectable *E. coli* or

thermotolerant coliforms (TTC) in drinking water samples [15]; in addition, they found that at least 52% of the general Sub-Saharan African population was exposed to fecally contaminated drinking water.

Wolf et al.'s (2014) systematic review analyzed the health impact of various drinking water and sanitation interventions, such as improved community level drinking water sources, piped water, higher quality piped water, point-of-use (POU) chlorine treatment, POU solar treatment and POU water filtration with and without safe storage. Despite the continuing risk presented by fecally contaminated drinking water sources, certain improvements in drinking water treatment and sanitation, particularly POU water filters, high quality piped water and sewer connections, were associated with substantial reductions in diarrheal disease [67]. POU water filters with safe water storage conferred the most protection [68]; they were associated with a 45% reduction of diarrheal disease over exposure to untreated unimproved water sources (RR=0.55, 95% CI: 0.38-0.81) and a 38% reduction of diarrheal disease over exposure to untreated improved water sources (RR=0.62, 95% CI: 0.42-0.93). Notably, POU solar and chlorine treatments did not significantly reduce diarrhea, after adjusting for bias from non-blinding [67].

Despite the fact that Rwanda is driving the most pronounced reduction of childhood mortality of any other country in the world, lack of access to improved water sources contributes substantially to diarrheal disease mortality in Rwanda and in the region as a whole [2,12]. Western Province, in particular, struggles with the lowest economic mobility and highest diarrheal disease prevalence in the country [12]. Leveraging this knowledge

and the existing evidence of various interventions, DelAgua Health initiated their *Tubeho Neza* program to distribute POU water filters with safe storage containers and improved cookstoves (as part of their goal to reduce childhood pneumonia) to 600,000 of the most socioeconomically disadvantaged households (*Ubudehe* levels 1 and 2) throughout Rwanda. The second phase of their work, which this baseline study references, will cover 100,000 of these households in Western Province as part of a cluster-randomized controlled trial of this paired intervention. This cross-sectional analysis of the baseline data will 1) ascertain balance of potential confounders between intervention and control households and 2) identify associations between household, socioeconomic and water, sanitation and hygiene indicators and 7-day period and point prevalence of diarrhea in children under 5 years-old.

Population of Western Province.

Education. Western Province has the highest percentage of men (17%) and women (26%) who have never obtained any formal education in Rwanda; however, this figure is largely influenced by age. Across the country, for example, 79% of women age 65 and over have no education, compared to 2% of girls between the ages of 10 and 14. We anticipate the same trend in our study population: the older our respondents are, the less likely they are to have had a formal education.

Access to improved water. The 2010 DHS indicates that 89.6% of households have access to improved sources of water (piped water, public tap/standpipe, borehole, protected dug well, protected spring, rainwater or bottled water) while another 7% have access to a

unimproved source (unprotected dug well, unprotected spring, tanker truck/cart with drum or surface water) [12]. While this appears promising, research from Phase 1 of the intervention indicated that among the 468 households our team assessed for total coliforms across 19 districts in Rwanda, 27.8% of "improved" water sources, 80.2% of unimproved water sources and 58.3% of stored water supplies exceeded the Rwanda Standard for Potable Water [13]. 34.7% of households do not treat their water; among households that do, boiling is the dominant method (58.5%) [12].

Access to improved sanitation. In rural areas, approximately 25% of households countrywide use unimproved sanitation facilities, the majority of which are pit latrines without a slab/open pit. Another 57% in rural areas have access to an improved/not shared pit latrine with a slab [12]. As these figures were calculated country-wide, we might expect the proportion of households that only have access to unimproved facilities to increase in the lower wealth quintiles, although these data were not disaggregated by socioeconomic status. In addition, only 10% of households have a place dedicated for handwashing, and among those households, only 21% actually have soap and water for hand washing. This figure decreases in Western Province, where only 4% of households have a designated handwashing place [12]. 77% of households in the lowest wealth quintile have no water, soap or cleansing agent available [12].

Methods

Study site and household selection. Western Province is located in western Rwanda, bordered by the Democratic Republic of the Congo and Lake Kivu to the west, Burundi to

the south, Southern Province (Rwanda) to the east/southeast and Northern Province (Rwanda) to the east/northeast. The province is divided into 7 districts: Karongi, Rutsiro, Rubavu, Nyabihu, Ngororero, Rusizi and Nyamasheke. These 7 districts are further divided into 96 administrative sectors, which in turn are divided into 538 cells and 3612 villages.

The paired intervention was randomized by sector at a 3:1 ratio (72 intervention sectors and 24 control sectors), and the intervention was distributed between September and December 2014. Control villages are scheduled to receive the intervention starting in March 2016, at the end of the follow-up period of this study. Our study team selected 87 intervention and 87 control study villages at random using population-proportional selection (PPS). Our sampling frame was the village-specific intervention distribution list of the *Tubeho Neza* program, which was derived from the Rwandan government's 2012 *Ubudehe* List. Households were considered eligible for the intervention if they fell into the lowest socioeconomic tertile, designated as *Ubudehe* levels 1 and 2. In addition to this socioeconomic requirement, households had to have at least one child under 5 years-old and have a primary cook 16 years-old or older to be eligible for selection into our study. Households in highly urbanized areas, such as Gisenyi in Rubavu and Kamembe in Rusizi, were excluded, as their use of coal over wood as cooking fuel precluded them from receiving the wood-burning cookstove distributed by *Tubeho Neza* as part of this paired intervention.

Within each of the 174 study villages, all eligible households were visited in random order and invited to participate in the baseline study and up to four follow-up rounds of the study

for up to 16 months. Eligible households were recruited from each village until either 10 households were enrolled or the eligible population was exhausted. Informed consent for the baseline study was obtained from the primary cook over of the household on behalf of all household members, and the primary cooked served as the principal respondent for the survey and all subsequent visits throughout follow-up. If a new primary cook was identified during subsequent visits, the consent procedure was repeated. Households were not provided with any incentive to participate, and their receipt of the *Tubeho Neza* intervention was not contingent upon their participation.

Data collection. Participating households of the overall study were visited once during baseline. Two households from each village were visited on three consecutive days as part of an intensive household air pollution study, but the primary survey in these households was conducted on the first day. Data on household demographics, primary cook and household head education, water source, use and handling practices, sanitation characteristics, hand-washing behaviors and other potential covariates were collected during the baseline survey. All survey instruments were first written in English and then underwent a double forward- and backward-translation process to obtain our final survey instruments in Kinyarwanda. All surveys were piloted before use in our study, and pilot participants and survey enumerators were asked to provide their feedback on the comprehensibility of the questions asked. Survey items were a combination of enumerator observations and participant questions; they were predominantly binary or multiple-choice, although some open-ended questions were asked where appropriate.

During the household visits, enumerators asked primary cooks about current and 7-day binary recall of diarrhea for themselves and for each child under 5 years-old in the household, using the local word for the condition, *impiswi*. Primary cooks were then asked about 7-day recall of diarrhea in each child under-5 according to the WHO case definition, which is three or more loose stools in any given 24-hour period. In addition, respondents were asked to provide information on duration of each reported illness (in days; care-seeking behavior for each illness from a CHW or health center; and vaccines received, as confirmed from a vaccination card administered by the Ministry of Health. If current diarrhea in any child under 5 years-old was reported, the child was assessed for 1) dehydration, using the WHO's Integrated Management of Childhood Illness (IMCI) criteria; 2) persistent diarrhea, defined as diarrhea lasting 14 days or more; and 3) dysentery, defined as blood in the stool at any time during that particular diarrhea episode [69]. Diarrhea with varying levels of dehydration was classified based on identification of a child with sunken eyes, IMCI criteria for either lethargy/unconsciousness or restlessness/irritability, degree of thirst and a skin pinch test on the child's abdomen. ALRI is also diagnosed in the field using an IMCI index incorporating the presence or absence of rapid breathing, chest indrawing and stridor [69]. All children with IMCI-classified general danger signs were referred into care through one of the village's designated CHWs. Regardless of the presence of diarrheal disease, middle-upper arm circumference (MUAC) measurements were obtained to rapidly assess nutritional status. Enumerators were trained on the IMCI protocol specific to the Rwandan Ministry of Health by the Ministry's own CHW trainers and pediatric clinic staff, the Rwandan Integrated Case Management of Childhood Illness [70].

In addition to point and period prevalence of diarrheal disease and respiratory disease, our team collected water samples at each participating household in sterile plastic WhirlPak bags (Nasco, Fort Atkinson, WI) containing one sodium thiosulfate tablet to neutralize any halogen-based disinfectants. Thermotolerant coliform (TTC) counts were obtained for each sample daily by trained laboratory staff using membrane filtration procedures prescribed for the Oxfam-DelAgua Incubator Kit (DelAgua, Wiltshire, UK). Using the kit's included pad dispenser, one membrane filter pad was placed on one metal petri dish for each water sample. Approximately 1.75mL of membrane lauryl sulfate medium (Oxioid Limited, Basingstoke, Hampshire, UK) was applied to fully cover each pad. Once the water sample was thoroughly mixed, the membrane filter was placed on the filter support of the filtration apparatus with sterile tweezers. 50-100mL of sample water was poured from the sample bag into the filtration funnel and vacuumed through the filtration funnel using a hand vacuum pump. Plates were incubated at 44°C for 16 hours prior to colony counting.

Population stratification and hierarchical selection. Data were collected in a hierarchical manner. The intervention was ultimately randomized by sector, and data were collected within multiple intermediate levels of observation:

Cluster: Sectors, which were randomized at a 3:1 ratio. Sectors are government administration areas; there are a total of 96 sectors in Western Province that lie within 7 districts. Cluster-level effects are mostly relevant for follow-up studies; cluster assignment

is accounted for in this cross-sectional baseline assessment through the sample weights assigned to households and individual children.

Village: There are 174 villages selected into this study, randomly selected using population proportional selection at a 1:1 intervention-to-control ratio. Village size will be considered as it may contribute to transmission of respiratory and diarrheal disease.

Household: 1582 households were ultimately enrolled across our study, for an average village enrollment size of 9.15 households per village. Potentially relevant data for our analysis at this level include indicators of socioeconomic status and water, sanitation and hygiene factors.

Individual: 2179 children were enrolled across our study, for an average of 1.38 children under 5 years-old enrolled per household. Children over 4 years-old at baseline were only enrolled if a younger child also lived in the house to avoid having whole households age out of the cohort throughout the anticipated 12 month follow-up period. Outcome-level data will be collected at this level, in addition to potential confounders such as age, gender and immunization status. Ensuing follow-up analyses will account for treatment-level effects; however, due to the cross-sectional nature of this baseline analysis, clustering will be accounted for within the village and household levels.

Statistical analysis. In order to establish that the selected households in the intervention and control study arms were balanced along potential confounders, descriptive statistics

were obtained for all demographic and socioeconomic factors and water, sanitation and hygiene (WASH) indicators at the household level incorporating household weights using SUDAAN (Research Triangle Institute, Research Triangle Park, NC USA).

Baseline prevalence of 7-day reported diarrhea, current diarrhea and current diarrhea with dehydration will be compared between intervention and control groups as an additional measure to ascertain balance between the two study arms at the primary outcome level. In order to assess the magnitude of association between household, sociodemographic and WASH factors and our primary diarrhea outcomes, adjusted multi-level random effects logistic regression models with robust standard errors to account for between-village variation. All models will be run along with calculated predicted margins, or the predicted response of all observations within each exposure level [71,72], in order to calculate risk ratios. Given population weights to be applied to each child in this complex survey dataset, the predicted margin is given by:

$$PM_r = \sum_{i=1}^{n} \delta_i w_i g(r, z_i, \hat{\theta}) / \sum_{i=1}^{n} \delta_i w_i \ ,$$

where δ_i=1 is the *i*th observation is in the sampled sub-population and 0 otherwise, w_i is the sample weight, *r* is the treatment group, z_i represents the values of the covariates and θ-hat is the sample-weighted estimator of the parameter vector [71]. As our study households are clustered within villages, unconditional variance of PM(r) will be estimated through Taylor linearization methods, allowing for unequal probability of selection of households within each village and household sampling with replacement. As all data are cluster-correlated, analyses were performed with SAS-Callable SUDAAN Ver. 11.0.1 (Research Triangle Institute, Research Triangle Park, NC).

Seven-day period prevalence of diarrheal disease. Model selection procedures were applied using methods outlined by Hosmer, Lemeshow & Sturdivant (2013) [73]. Variables selected for model inclusion underwent prior bivariate analyses with our outcome of interest, 7-day period prevalence of diarrheal disease. CMH chi-square tests were performed and categorical variables were selected for initial inclusion into the multivariate models if they were even mildly significant at this bivariate level ($p \geq 0.4$). Based on these procedures, the following variables were excluded from consideration: water treatment (wu4), number of household members per sleeping room (pph), designated handwashing location (hw1), water available for handwashing (hw3), number of households sharing toilet (tf6), method of water treatment (wu5), frequency of water treatment (wu6), MMR (cv5) and rotavirus (cv6) vaccine.

Backwards selection was first performed on a full model specifying all potential covariates. Each reduced model was independently assessed for confounding and interaction. A confounding assessment was then performed on all potential covariates, using the following full models to compare each partial model against:

$$logit(p_1) = \ln\left(\frac{p_1}{1-p_1}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(tf4) + \beta_4(tf5) + \beta_5(tf5a) + \beta_6(SES) + \beta_7(cd3) + \beta_8(cd6) + \beta_9(wu3rt) + \beta_{10}(wu4) + \beta_{11}(hw1) + \varepsilon$$

$$logit(p_2) = \ln\left(\frac{p_2}{1-p_2}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(tf4) + \beta_4(tf5) + \beta_5(tf5a) + \beta_6(SES) + \beta_7(cd3) + \beta_8(cd6) + \beta_9(wu3rt) + \beta_{10}(wu4) + \beta_{11}(hw1) + \varepsilon$$

Where p_1 is the probability of a child having respondent-defined diarrhea in the last 7 days, p_2 is the probability of a child having diarrhea fitting the WHO case definition in the last 7 days, tf2a=toilet type, wu1=water source, tf4=feces on floor within 1M of toilet, tf5=shared toilet, tf5a=location of toilet, SES=SES index (derived from PCA), cd3=gender, cd6=age (in months), wu3rt=round-trip time to fetch water (in minutes), wu4=drinking water is treated and hw1=designated handwashing location in household. Further results of the individual confounding assessments for each exposure of interest are described in Appendix 1.2.

Next, unweighted multivariate logistic regression analyses were performed with stepwise selection procedures in order to further refine the final model to be used to examine the relationship of water and sanitation factors and reported diarrheal disease. At each stage of the selection process, a covariate had to be significant at a significance level of 0.35; in order for each covariate to be eligible for final model inclusion, a significance level of 0.30 was required. Based on these criteria, final model specification is described further in the results section. Survey weights and robust standard errors were applied to the final model to account for within-village clustering of diarrhea cases.

Water quality. Two model techniques were considered to examine the association between various water and sanitation indicators and water quality, measured by thermotolerant

coliform counts: proportional odds and multinomial logit regression. Proportional odds models will be fit with the SUDAAN MULTILOG procedure with a cumulative logit link; multinomial logit regression models will also be fit with the SUDAAN MULTILOG procedure, but with a generalized logit link. For the proportional odds models, the probability of worsening water quality (higher TTC counts) across levels of various exposure variables will be modeled. In the event the proportional odds assumption is not met, the multinomial logit model will be used to model the probability of a water quality sample being in any of the last three WHO JMP exposure categories compared to the lowest WHO JMP exposure category (TTC <1 CFU/100mL). Regardless of the model chosen, generalized estimating equations model-fitting techniques will be run under an independent working assumption and a Zeger robust variance estimator.

In order to assess the association between household, water and sanitation factors and water quality, the dependent variable was an ordinal variable created to reflect WHO JMP risk categories based off of the TTC counts obtained from the water samples in the study. Backwards selection procedures were also used to determine the optimal model specification for water quality. The full model used for all reduced model comparisons was:

$$logit(p_3) = \ln\left(\frac{p_3}{1-p_3}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(tf4) + \beta_4(tf5) + \beta_5(tf5a) + \beta_6(SES) + \beta_7(cd3) + \beta_8(pph2) + \beta_9(wu3rt) + \beta_{10}(wu4) + \beta_{11}(hw1) + \varepsilon$$

Where p_3 is the probability of any given water sample being at any WHO-defined water quality risk category and pph2 reflects the total number of household members per sleeping room in the house.

Data adjustments. In order to accurately reflect disease burden, seven-day period prevalence of diarrhea was adjusted if a negative response was recorded for "Has your child had diarrhea in the last 7 days" but a positive response was recorded for "Does your child currently have diarrhea?" Child age in months was checked against the recorded age in months. Age in months was rounded up after 15 days, and all age values were checked and altered if they did not reflect this. If date of birth was not available for a child, then the child's total age in months was checked against the caregiver's reported age for the child. If only a child's age in years was known, then the child's age in months was calculated by multiplying the total age in years by 12 (e.g., if a child was reported as being 3 years-old, then the age was recorded as 36 months). The total number of children under-5 living in the household was also checked against the total number of children under-5 recorded in the survey, as enumerators should have entered in each child under-5 years-old. If the number of children initially reported to be living in the household did not match the number of children actually entered in the survey, then the reported number of children under-5 was changed to reflect the number of children recorded in the survey. Total number of household members was also changed to reflect this change in the number of children under-5 currently living in the household.

The water source variable (wu1) contained three levels for which no cases of diarrheal disease were reported: “water source: piped water into dwelling,” “water source: piped water into yard/plot” and “water source: protected dug well.” As these three levels of the water source covariate perfectly predicted diarrheal disease outcomes, they created model convergence problems. For this reason, they were excluded from all multivariate analyses, and the following levels were retained: “Public tap/standpipe,” “Hand pump (borehole),” “Protected spring,” “Rainwater,” “Unprotected dug well,” “Unprotected spring,” “pond/lake” and “river/stream.”

Ethical approval. The study protocol, survey instruments and informed consent was reviewed and approved by the Emory University Institutional Review Board (Ref #: 73615), the London School of Hygiene and Tropical Medicine Research Ethics Committee (Ref # 7711), the Rwandan National Ethics Committee (Ref # 1497) and the National Health Research Committee of Rwanda (Ref # NHRC/2014/PROT/0163).

Results

A total of 1582 households containing 2179 children under 5 years-old were enrolled during the baseline survey. Among the children enrolled, 1077 (50.12%) were female and 1072 (49.88%) were male. 212 (9.73%) of children were 0-5 months-old, 270 (12.40%) of children were 6-11 months-old, 458 (21.03%) of children were 12-23 months-old, 489 (22.45%) of children were 24-35 months-old, 541 (24.84%) were 36-48 months-old, and 208 (9.55%) were 48-60 months-old. The mean and median age of the respondent, who

was the primary cook of the household, was 32.7 and 31 years-old, respectively, and ranged from 16 to 91. While an average of 1.45 children per household was recorded, an average of 1.38 children per household were enrolled in the study. Selected baseline characteristics for all households, representing potential confounders, are present in Table 5.1, which reveals that the intervention and control arms were well-balanced at baseline on all potential confounders. For the PCA used to derive the socioeconomic index, approximately 27.06% of the total variance exhibited by education, asset and housing construction variables was explained by the 1st principal component. Factor loading scores are presented in Appendix 1.

Seven-day period and point prevalence of diarrhea, stratified by treatment arm, are presented in Table 5.2. Both period and point prevalence of diarrhea appeared to be higher in the intervention arm (16.11%, 95%CI: 13.85-18.66) than in the control arm (13.46%, 95%CI: 11.51-15.68), although this difference was not statistically significant (F=1.55, p=0.22). Period and point prevalence of diarrhea also appeared to be higher in younger age groups and in males. Out of the total 2179 children enrolled in the study, there were 320 children for whom diarrhea was reported in the last 7 days; therefore, the total weighted period prevalence of diarrhea was 15.39% (95% CI: 13.64-17.33). Eighty-two children had caregiver-reported diarrhea on the day of the survey, resulting in a total weighted point prevalence of diarrhea of 3.91% (95% CI: 3.08-4.96%).

Tables 5.3a and 5.3b indicate that 7-day diarrhea prevalence appears to be affected by various household water and sanitation factors. After fitting the model, 343 children were

eliminated from the analysis if data were missing for one of the covariates in the model. Overall, out of 2179 children enrolled at baseline, 1836 were included in the multivariate analyses outlined above. We make the assumption that the observations for these 343 children are missing at random, so all logistic regression analyses can be generalized to all children under 5 years-old in our study area. Out of the 174 clusters included in the analysis, 172 were used to fit the model, as 2 clusters did not have a child for which non-missing data across all covariates were available. The minimum cluster size was 1, while the maximum cluster size was 22. Overall, 294 out of the 1836 children used to fit the model had caregiver-reported diarrhea in the last 7-days, while 275 children had caregiver-reported diarrhea in the last 7-days that fit the WHO case definition. The 28 df Wald *F*-test indicates that we can reject the null hypothesis that all regression coefficients except for the intercept are equal to zero ($F=3.901$, $p<0.001$). Wald F-tests of the main effects indicate that water source, toilet type, age and the interaction between water source and round-trip time to water source are statistically significant ($p<0.05$). In addition, variables for feces within 1M of toilet ($F=3.05$, $df=1$, $p=0.082$) and shared toilet ($F=3.20$, $df=1$, $p=0.076$) approach significance. Children in households with the following characteristics appear to have higher odds of having diarrhea in the last 7 days: household with a composting toilet; water source=public tap or standpipe or pond/lake; use of a shared toilet and low socioeconomic status (1^{st} and 2^{nd} quintile groups). Age also appeared to be a major predictor; the odds of diarrhea in the last 7 days were higher in the youngest age groups, and tapered in a linear fashion by age.

1. Primary water source

1.1. Respondent-defined seven-day prevalence of diarrhea

The weighted model-adjusted risk of 7-day period prevalence of diarrhea, applying the WHO case definition,is given by predicted marginal prevalence. Water sources were categorized by WHO criteria for "improved" vs. "unimproved" water sources. No significant disadvantage was conferred among children in households obtaining their drinking water from either unimproved water sources (aRR=0.99, 95%CI: 0.72-1.36) or surface water sources (aRR=1.25, 95%CI: 0.79-1.98) compared to children in households obtaining their drinking water from improved drinking water sources (Table 5.3a).

When further disaggregated into water source categories, the reference group for all analyses was protected spring, based on a combination of low diarrhea prevalence (crude proportion=13.42%, 95%CI: 11.23-15.98)) and larger population representation (n=1224) relative to other water source categories. Relative to households obtaining their water from a protected spring, children in households obtaining their water from a public tap/standpipe (aRR=1.62, 95%CI: 1.18-2.22), a hand pump or borehole (aRR=2.20, 1.52-3.18), rainwater (aRR=3.90, 95%CI: 2.61-5.85) and a pond or lake (RR=2.10, 95%CI:1.22-3.60) had an increased risk of reported diarrhea in the last 7 days (Supplemental Table 5.3a)). Notably, no diarrhea cases were reported among children in the 23 households obtaining their drinking water from piped water into the dwelling, piped water into the yard/plot or a protected dug well (Supplemental Table 5.3a); because these levels of the water source variable fully separated case data, maximum likelihood estimates could not be estimated and prevalence ratios are not displayed.

No significant relationship was observed for round-trip to and from water source, although it did significantly modify the effect of water source type on seven-day prevalence of diarrhea ($F=5.270$, $p<0.0001$) in unweighted models. While not significant, the model-adjusted risk for diarrhea seemed to increase as round-trip time to water source increased. In addition, no significant relationship was observed between diarrhea and water quality, measured by TTC counts in CFUs/100mL, as the risk of diarrhea appeared to be the same across all WHO-defined water quality risk categories (Tables 5.5a and 5.5b).

1.2 Seven-day prevalence of diarrhea, adhering to WHO case definition for diarrhea

Similar to the results presented for respondent-defined seven-day prevalence of diarrhea, no significant disadvantage was conferred among children in households accessing their water from unimproved sources (aRR=0.89, 95%CI: 0.62-1.27) or from surface water (aRR=1.40, 95%CI: 0.91-2.14) compared to households obtaining their water from improved sources.

Despite this, further disaggregating improved and unimproved water sources into individual water source types yielded varying results. Relative to households obtaining their water from a protected spring, children in households obtaining their water from a public tap/standpipe were 74% more likely to have had respondent-reported diarrhea fitting the WHO case definition in the last 7 days (aRR=1.74, 95%CI: 1.30-2.33). As opposed to children with respondent-defined diarrhea in the last 7 days, application of the WHO case definition diminished the association previously observed between 7-day period prevalence of diarrhea and obtaining water from a hand pump/borehole (aRR=1.40,

95%CI: 0.70-2.77). Additionally, collecting water from rainwater (RR=3.56, 95%CI: 2.42-5.25) and a pond or lake (aRR=1.96, 95%CI: 1.02-3.78) presented significant increased risk over collecting water from a protected spring (Supplemental Table 5.5b).

No significant relationship was observed for round-trip to and from water source and 7-day prevalence of WHO-defined diarrhea. Model-adjusted risk for diarrhea appeared to be nearly identical between households with >30 minute round-trip time to fetch water and households with water close by (0 to 15 minute round-trip time). In addition, there was no significant association between water quality and WHO-defined diarrhea in the last 7 days. No apparent trend, significant or not, is present for either roundtrip time to water source or water quality (Table 5.3b).

1.3. Water source and water quality

In general, surface water (aRR=0.25. 95%CI: 0.07-0.88), unimproved water sources (aRR=0.67, 95%CI: 0.43-1.03) and piped water (aRR=0.08, 95%CI: 0.01-0.72) appeared to be less likely than other improved water sources to be in compliance with WHO standards (<1 TTC CFU/100mL). In addition, surface water and unimproved water appeared to be more likely to be designated as high risk (>100 CFU/100mL) compared to improved, non-piped water sources (aRR=2.00, 95%CI: 1.45-2.76; aRR=1.40, 95%CI: 1.00-1.97, respectively).

2. Toilet type and sanitation characteristics

2.1. Seven-day prevalence of diarrhea

The Wald F-test assessing the effect of toilet type on 7-day prevalence of diarrhea was significant (F=35.06, p<0.0001). Too few open defecation households were in our study area to include in this analysis. Unimproved and shared improved forms of sanitation did not appear to pose any substantial excess risk or benefit (Table 5.3a). After disaggregating by toilet type, children in households with composting toilets (aRR=1.91, 95%CI: 1.13-3.23) and with no toilet access (aRR=3.54, 95%CI: 2.32-5.40) appeared have a greater risk of having had diarrhea in the previous 7 days than children living in households with a pit latrine with no slab. Living in a house with a ventilated pit latrine or a pit latrine with a slab presented no significant risk or benefit (Supplemental Table 5.3a).

Toilet location (within house, within plot, within compound and outside of compound) was not significantly associated with 7-day prevalence of diarrhea. Fecal matter within 1 meter of the toilet was also not associated with diarrhea (aRR=1.18, 95%CI: 0.93-1.51). Shared toilet facilities was positively associated with 7-day prevalence of diarrhea (aRR=1.53, 95%CI: 1.12-2.09), although the number of households sharing a toilet did not appear to affect diarrhea period prevalence. 92.7% of households that shared a toilet shared it with only 1 or 2 other households, reflecting high population homogeneity on this factor.

2.2. Seven-day prevalence of diarrhea, adhering to WHO case definition for diarrhea

The Wald F-test assessing the effect of toilet type on 7-day prevalence of diarrhea adhering to the WHO case definition was significant (F=92.10, df=5, p<0.001). The significant associations between toilet type and 7-day prevalence of diarrhea adhering to the WHO definition were similar to those identified for 7-day prevalence of diarrhea with the primary

cook's own definition. Among the WHO sanitation ladder categories, no associations with diarrhea were evident, and too few open defecation households were present to incorporate into any analyses (Table 5.3b). After disaggregating the sanitation variable into its constituent sanitation types, children living in households with composting toilets were nearly twice as likely to have had diarrhea in the last 7 days as children living in households with a pit latrine with no slab (RR=1.81, 95% CI: 1.08-3.04). Out of the 35 children living in households with non-shared composting toilets, 11 diarrhea cases were identified (aRR=2.49, 95%CI: 1.48-4.17); however, among the 15 children living in households with shared composting toilets, only one diarrhea case was identified. Living in a household with a pit latrine with a slab or a ventilated pit latrine presented no significant risk or benefit (Table 5.3).

Toilet location was not significantly associated with 7-day prevalence of WHO-defined diarrhea. Children living in households where fecal matter was observed within 1 meter of the toilet were 35% more likely to have had diarrhea adhering to the WHO case definition in the previous 7 days (RR=1.35, 95%CI: 1.02-1.80). Shared toilet facilities were also significantly associated with diarrhea under the WHO case definition (RR=1.54, 95%CI: 1.13-2.10) (Table 5.3b).

2.3. Toilet type and sanitation characteristics and water quality

Toilet type did not appear to be associated with water quality. Due to stratification of the outcome variable into four levels according to prescribed WHO water quality standards, strata-specific samples sizes were too low to disaggregate the sanitation variable into its

component toilet types (Table 5.6). There was no significant association or trend between toilet location and water quality. Similarly, toilet area cleanliness (feces within 1M of toilet) and sharing a toilet with one or more households did not appear to significantly affect water quality. Households with a designated handwashing station had a higher proportion of water samples in compliance with WHO standards (aRR=2.42, 95%CI: 1.47-3.98) (Table 5.6), and households with handwashing locations appeared to have fewer water samples designated as "high risk" (aRR=0.51, 95%CI: 0.19-1.33).

3. Household characteristics

3.1. Seven-day prevalence of diarrhea.

While household crowding, measured by number of persons per sleeping room, did not appear to be associated with diarrhea, households in the lowest socioeconomic tier had a higher risk of diarrhea (aRR=1.58, 95%CI: 112-2.22) compared to households in the highest socioeconomic tier. Diarrhea prevalence mostly appeared to decrease as socioeconomic status improved (Table 5.3a).

3.2. Seven-day prevalence of diarrhea, adhering to WHO case definition for diarrhea.

Similarly, households in the lowest socioeconomic tier appeared to have an elevated risk of diarrhea fitting the WHO case definition (aRR=1.83, 95%CI: 1.27-2.65). Risk generally tended to decrease as socioeconomic status improved (Table 5.3b).

4. Child age

4.1. Seven-day prevalence of diarrhea

Compared to children who were 48-60 months-old at baseline, children in younger age groups had a higher risk of diarrhea. Children who were 6-11 months-old had the highest risk (RR=3.63, 95%CI: 2.05-6.41) and gradually decreased as age increased.

4.2. Seven-day prevalence of diarrhea, adhering to WHO case definition for diarrhea

Age-specific prevalence of diarrhea adhering to the WHO case definition fit a similar pattern. Children who were 6-11 months-old at baseline had the highest risk of diarrhea compared to children who were 48-60 months-old (RR=3.28, 95%CI: 1.74-6.17). After that point, diarrhea prevalence gradually decreased as age increased.

5. *Water quality and diarrhea.*

Overall, 336 (25.74%, SE=1.77) of drinking water samples were free of TTC while 484 (35.89%, SE=2.04) of water samples contained ≥ 100 CFU/100 mL, designated as "high risk" by the WHO's Joint Monitoring Programme (Figure 5.2). At baseline, the two treatment arms did not appear to substantially differ in terms of overall levels of water contamination, as measured by TTC counts.

Discussion

This cross-sectional analysis characterizes population characteristics from the baseline data of the largest randomized controlled trials on point-of-use water filters and improved cookstoves to date. Populations assigned to both the intervention and control arms appeared to be well-balanced on all potential confounders at baseline. Seven-day prevalence of

diarrhea appeared to be moderately higher in the intervention group (16.11%) than the control group (13.46%) at baseline, which is a difference that appears to be driven principally by the youngest children and should be accounted for when assessing the overall impact of the intervention post-distribution on diarrheal disease. Diarrhea prevalence was higher in children under 12 months of age in the intervention group (20.60%, 95%CI: 15.26-27.20) compared to the control group (13.91%, 95%CI: 10.07-18.91). This population already tends to be disproportionately affected by diarrhea due to high diarrhea prevalence that coincides with the post-weaning phase. This relationship between age and diarrhea prevalence in this study indicates that children in Western Province Rwanda who are 0 to 5 months-old have a 2.5-fold risk of diarrhea (RR=2.53, 95%CI: 1.26-5.07) and that children ages 6 to 11 months have nearly quadruple the risk of diarrhea (RR=3.63, 95%CI: 2.05-6.41) compared to children in the oldest age group who were 48 to 60 months-old. This age distribution of diarrhea is in line with other studies. For example, Fawzy et al. (2011), in a randomized controlled trial to assess the impact of short vs. long duration of breastfeeding of children born to HIV-positive mothers in Zambia, found that diarrhea prevalence markedly increased among children 6 to 15 months-old and declined again among children 24 months-old and older. They also found that children who had weaned had over twice the risk of diarrhea compared to children who were still breastfeeding [74]. In a systematic review, Fischer-Walker et al. (2012) also found that diarrhea incidence peaked in children 6-11 months-old in Sub-Saharan Africa, and they recommended exclusive breastfeeding through the first 6 months of life and continued breastfeeding through 24 months to alleviate this burden among children in this age group [75]. Given that seven-day prevalence of diarrheal disease surged shortly after

the predicted time that exclusive breastfeeding ends, as determined from the Rwanda 2010 DHS (Figure 5.3), this surge in diarrheal disease risk in children who are 6 to 11 months-old in our study is likely also attributed to high exposure to novel pathogens in the immediate post-weaning stage.

Water source characteristics

Obtaining drinking water from an "improved" water source did not appear to affect diarrhea prevalence, although certain improved water sources appeared to be more preventive than others. While only 11 households out of the 1582 enrolled in the study obtained water primarily from piped water into the home, yard or plot, no diarrhea cases were reported in the previous 7 days among the 14 children living in these households. There were also no diarrhea cases among another 14 children living in 10 households that obtained their drinking water from a protected dug well. This does indicate a potential substantial protective benefit of having piped water in close proximity to the home or access to a protected well. Notably, piped water in close proximity to the home or obtaining water from a protected well did not appear to be associated with water quality, measured by TTC in CFU/100mL. Interestingly, obtaining water primarily from a public tap or standpipe was associated with a 74% and 62% increase in reported respondent- and WHO-defined seven-day period prevalence of diarrhea, respectively. In addition, obtaining water from a hand pump or borehole was associated with over a 2-fold increase in WHO-defined seven day period prevalence of diarrhea (aRR=2.20, 95%CI: 1.52-3.18) compared to children in household collecting their water from a protected spring. While the majority of our population lives in rural villages and dispersed housing, it may be important to adjust

the data for household density, as standpipe construction is resource-intensive and may be prioritized for more densely population areas [76]. Both standpipes and boreholes may either consolidate contamination from multiple households or serve as a proxy measure for some other risk factor, like household density itself. As both standpipes and boreholes are considered "improved" water sources by the WHO JMP criteria [60], this data does not comport with previous synthesis of data on the association between water supply and diarrhea (Clasen et al., 2009), which generally provides evidence of a protective benefit conferred from these two source types. Additionally, water obtained from a hand pump or borehole appeared more likely to be free of fecal coliforms compared to water collected from a protected spring, so diarrhea among children in these households could be largely attributed to other environmental exposures or to inadequate water transportation or storage prior to the survey day.

Surface water is considered an "unimproved" water source [60] and 20 households with 30 children collected their drinking water primarily from a pond or lake and 68 households with 90 children collected their drinking water primarily from a stream or river. Collecting water from a pond or lake was associated with double the risk of respondent-defined and WHO-defined diarrhea episodes, while no association was observed among children in households collecting their water from a stream or river compared to children in households obtaining their water from a protected spring. Households obtaining water from both surface water categories were significantly less likely to have a water sample fitting the WHO standard of TTC at <1 CFU/100mL and were significantly more likely to have a

"high risk" water sample of TTC at more than 100 CFU/100mL, indicating that the water source was contributing significantly to drinking water contamination.

Round-trip time to and from the water source varied by the water source type itself. Generally, households obtaining their water from an unprotected or a protected spring appeared to be more likely to spend 16 or more minutes collecting water, while households obtaining their water from a public tap or standpipe seemed more likely to reside within a 5 minute round trip to their water source (RR=0.87, 95%CI: 0.82-0.93) than households obtaining their water from a protected spring. Time to water source itself was not significantly associated with respondent- or WHO-defined 7-day prevalence of diarrhea.

Respondent reporting of household drinking water treatment practices was not associated with diarrhea and revealed no evidence of confounding. It is likely that current household drinking water practices are either not effective or consistent enough to yield any substantial reduction in diarrheal disease risk. While the presence of a specific handwashing location was not associated with diarrhea prevalence, households with a handwashing location were significantly more likely to have a water sample free of TTC contamination (aRR=2.44, 95%CI: 1.47-3.98) and appeared to be about half as likely to yield highly contaminated samples (aRR=0.51, 95%CI: 0.19-1.33) compared to households with no handwashing location.

Sanitation characteristics

Children in households using shared sanitation facilities were 53% more likely to have had respondent-defined diarrhea in the last 7 days and 54% more likely to have had WHO-defined diarrhea in the last 7 days, after controlling for water source, toilet location in relation to the household and age of the child. As opposed to findings from Fuller et al. (2014) [78], socioeconomic status appeared to neither modify nor confound this effect. This association stood, regardless of the number of households sharing any given toilet, household size and toilet type, which were considered as potential confounders in this analysis but did not affect effect size estimates. This comports with a synthesis of data on shared sanitation and diarrhea in 51 countries, which demonstrated an anticipated 32% increase in diarrhea in households that shared sanitation facilities in Rwanda, based off of DHS data [78]. This also aligns with a separate meta-analysis of 11 studies in 9 countries that found that shared sanitation resulted in a 44% increase in the odds of diarrhea incidence, even though there was substantial study heterogeneity [79]. Despite these associations, it does not appear that shared sanitation is linked to household drinking water contamination, as shared sanitation is not associated with either ideal or high-risk water quality.

Toilet area cleanliness, assessed by each enumerator as having observed fecal matter within 1 meter of the toilet, was not associated with respondent-defined diarrhea, but was associated with a 35% increase in WHO-defined diarrhea in the last seven days. This indicates that toilet area cleanliness may affect fecal contamination in the broader environment or affect diarrhea risk by coming into contact with fecal matter itself. Toilet area cleanliness was not associated with ideal or high-risk water quality.

Compared to having a pit latrine with no slab, having a pit latrine with a slab was not associated with respondent- or WHO-defined seven-day period prevalence of diarrhea. In addition, having a pit latrine with a slab was not associated with ideal or high-risk water quality. Generally, this aligns with previous research indicating that the presence of a slab in a pit latrine does not drive any meaningful difference in fecal indicator bacteria levels found in the household environment [80].

Composting toilets nearly doubled the risk of both respondent- and WHO-defined diarrhea, despite being considered an "improved" form of sanitation [60]. There is considerably geographic heterogeneity regarding the distribution of composting toilets and the association between composting toilets and diarrhea. While only 50 households have composting toilets in this study, 82% of households with composting toilets are in the three northernmost districts of Western Province (Rubavu, Rutsiro and Nyabihu), where the volcanic soil may be too rocky to dig a deep pit for a latrine. A recent modeling exercise found that upgrades to "improved" sanitation, including composting toilets, do not sufficiently remove pollutants and pathogens from the environment, and that more efforts need to be directed towards safe storage, movement and treatment of feces [81]. Specific to northern Rwanda, knowledge of composting toilets is generally poor; after assessing knowledge and proper use of household and community composting toilets among local authorities and community members, implementers of a composting toilet intervention found that fecal matter was often improperly transported and improperly disposed of in open pits and that communities lacked designated composting sites [82]. Notably, only

non-shared composting toilets posed excessive risk; shared composting toilets, like the EcoSan toilets observed at the community-level, do appear to be protective. The increased risk associated with composting toilets in our study may not be a result from the toilets themselves, but the improper disposal, transport and use of fecal material at the household-level.

The presence of a designated handwashing location in the home was not significantly associated with diarrhea, nor was it identified as a confounder in assessing the relationship between toilet characteristics and diarrhea. Of the less than 3% of households that had a designated handwashing location, about half had water designated for handwashing and only about 14% had soap, indicating that handwashing was likely to unusual and infrequent to influence incidence of diarrheal disease.

Water quality

There was no significant association between water quality, assigned to standard WHO risk categories, and either respondent- or WHO-defined 7-day prevalence of diarrhea. As thermotolerant coliforms only weakly co-occur with enteric pathogens, these results were not unanticipated as they align with other observational study assessing the relationship between coliform counts and diarrheal disease. A systematic review by Gundry et al. (2004) assessed 16 observational studies relating *E. coli* and TTC microbial indicators at point-of-use to diarrheal disease in pre-school children and determined that there was no

significant relationship between these indicator bacteria and diarrheal disease. Intervention studies addressing point-of-use drinking water illustrate that contaminated drinking water is linked to diarrheal disease [67]; the fact that we found no association between drinking water quality and diarrheal disease is likely due to the method of water quality ascertainment than the level of contamination itself, as fecal indicators are often too variable and nonspecific to truly characterize exposure to disease-causing pathogens at any given time [84]. Furthermore, our study collected water samples on the survey day, but disease was determined retrospectively; therefore, no temporal link can be ascertained between consumption of contaminated drinking water and diarrheal disease. Further complicating this issue is that most households may have intermittently contaminated water samples [85]. In a study that prospectively assessed diarrhea 3 to 100 days after collecting monthly drinking water samples, children living in households with consistently highly contaminated water (>100 CFU of *E. coli* per mL) were significantly more likely to have diarrhea; in fact, each 10-fold increase in *E. coli* concentration in drinking water resulted in a 14% increase in diarrhea at the next visit [85]. Together, these results indicate that TTC counts themselves may not be particularly useful on an observational basis, and that prospective *E. coli* indicator measurements may better predict incidence of diarrheal disease.

There were several limitations associated with this study. First and foremost, diarrheal disease was ascertained by seven-day recall by the respondent, lending the potential for recall bias of our outcome measure. There is some potential for misclassification bias of water and sanitation characteristics, as categorization was based solely on each

enumerators' observations and participants' responses. Selection of households was based off of the Government of Rwanda's 2012 *Ubudehe* List, so our study did not capture residents who were on the list but had moved away, or residents who had newly arrived in the village but were not on the list. These residents may have differed along some fundamental characteristics from our study population, which may affect the generalizability of these results. This limitation also resulted in differing cluster sizes between villages, rather than the standard ten households per village we were originally seeking for this study. Finally, given that TTC counts appeared to be associated with various water and sanitation factors but not with diarrheal disease may indicate that a temporal relationship is necessary to establish a link between TTC data and disease.

While diarrheal disease remains one of the largest contributors to childhood mortality in Rwanda, novel point-of-use methods for household water treatment may contribute substantially to ameliorating this problem. The results reported here suggest that before our intervention was distributed, the following factors were associated with diarrhea: 1) drinking water from a public tap or standpipe, rainwater, borehole, pond/lake or river/stream; 2) shared sanitation; 3) toilet area cleanliness; and 4) other sanitation characteristics, such as composting toilets and open defecation. While open defecation and water samples taken from household drinking water obtained from boreholes were associated with low TTC contamination and water samples taken from household drinking water obtained from ponds, lakes, rivers and streams were associated with high TTC contamination, water quality data obtained from the study were not associated with 7-day prevalence of diarrhea. This study illustrates the characteristics of our RCT population as

they were enrolled in the baseline round, and demonstrates that despite these identified associations, our population was well-balanced across all potential confounders and exposures at baseline between intervention and control arms.

Appendix 5.1: Socioeconomic indicators and principal components analysis.

Variables included in the PCA included two education variables (primary cook education and household head education), seven asset variables (electricity access; possession of a radio, mobile telephone, mattress, agricultural land and cows; ownership of the home) and three housing construction variables (roof type, floor type, wall type). All sociodemographic household factors are listed in Table S5.1. The following variables were not excluded from consideration in the PCA due to low standard error (SD<1), high population homogeneity (>95%) and collinearity: ownership of a television, bicycle, motorcycle, boat and non-agricultural land. Other variables commonly included in PCAs for the purpose of obtaining a socioeconomic index, such as sanitation infrastructure, water source and household crowding, were not included since these were considered primary exposures of interest.

Based off of the factor loading scores present in Table A2, a continuous variable for socioeconomic status was generated in the following manner:

SES=(pced*0.47078)+(hhed*0.46429)+(SD4*0.68138)+(SD5A*0.57888)+(SD7A*0.739
96)
+(SD8*0.77252)+(SD13*0.16014)+(SD14*0.05156)+(SD15*0.19110)+(floor*0.76075)
+(wall*0.38565)-(roof*(0.14222))

This SES score was then split into quintiles to provide a SES quintile index with which to categorize our study households.

Appendix 5.2: Interaction and confounding assessment

Interaction was assessed by fitting the full model with all potential confounders with all potential interaction terms. The model was subject to backwards selection, where a variable would have to be significant at α=0.3 in order for it to remain in the model. Models were refit at each successive step, and variables had to be eligible at α=0.35 in order to remain in the model. All potential confounders and exposures remained in the model for the confounding assessment, but potential interaction terms were eliminated.

After assessing potential interaction terms for the association between all model covariates and diarrhea reported in the previous week (ch2), only the term for the interaction between water source (wu1) and round-trip time to water source (wu3rt) remained in the model (X^2=8.53, p=0.29).

After that, weighted logistic regression models with robust standard errors to account for clustering were fit for each exposure of interest. The full model was described as follows:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(tf4) + \beta_4(tf5) + \beta_5(tf5a) + \beta_6(SES) + \beta_7(cd3) + \beta_8(cd6) + \beta_9(wu3rt) + \beta_{10}(wu4) + \beta_{11}(hw1) + \omega(wu1 * wu3rt) + \varepsilon$$

Where tf2a=toilet type, wu1=water source, tf4=feces on floor within 1M of toilet, tf5=shared toilet, tf5a=location of toilet, SES=SES index, cd3=gender, cd6=age (in months), wu3rt=round-trip time to fetch water (in minutes), wu4=drinking water is treated and hw1=designated handwashing location in household.

Identical entry and retention criteria were used when assessing interaction in models fit to examine the association of these exposures and diarrhea fitting the WHO case definition in the last 7 days (ch3). After performing backwards selection on all model covariates and

potential interaction terms, no interaction term was significant at the α=0.30 level. The full model defined for all covariates and ch3 was as follows:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(tf4) + \beta_4(tf5) + \beta_5(tf5a) + \beta_6(SES) + \beta_7(cd3) + \beta_8(cd6) + \beta_9(wu3rt) + \beta_{10}(wu4) + \beta_{11}(hw1) + \varepsilon$$

Confounding was then assessed for all models. For each exposure of interest, the full model was first fit and predicted marginal risk ratios obtained for each level of each exposure of interest. Covariates were then removed one at a time and the model re-fit. A particular covariate was permanently removed from the model if it did not appear to substantially impact the risk ratio for all levels of a particular exposure; i.e., the ratio of risk ratios must not have differed by more than 10%.

Confounding assessment

1.1. Toilet type and 7-day diarrhea. After confounding assessment, sesind tf5a cd3 hw1 and wu4 were removed. This left a final model specified as:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(tf4) + \beta_3(wu3rt) + \beta_4(cd6) + \beta_5(tf5) + \varepsilon$$

For examining sanitation ladder categories (saladder), sesind tf5 tf5a wu3rt cd3 hw1 and wu4 were removed, leaving the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(saladder) + \beta_2(wu1) + \beta_3(cd6) + \beta_4(tf4) + \varepsilon$$

1.2. Toilet type and 7-day diarrhea fitting WHO case definition. The following variables were removed: tf5 tf5a cd3 hw1 and wu4. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(tf4) + \beta_6(SES) + \beta_8(cd6) + \beta_9(wu3rt) + \varepsilon$$

For examining sanitation ladder categories, tf5 tf5a wu3rt cd3 hw1 and wu4 were removed.

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(saladder) + \beta_2(wu1) + \beta_3(cd6) + \beta_4(tf4) + \beta_5(SES) + \varepsilon$$

2.1. Feces within 1M of toilet and 7-day diarrhea. The following variables were removed: tf5 tf2a sesind wu3rt tf5a cd3 hw1 and wu4. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(wu1) + \beta_2(tf4) + \beta_3(cd6) + \varepsilon$$

2.2. Feces within 1M of toilet and 7-day diarrhea fitting WHO case definition. The following variables were removed: tf5 tf2a sesind wu3rt tf5a cd3 hw1 and wu4. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(wu1) + \beta_2(tf4) + \beta_3(cd6) + \varepsilon$$

3.1 Shared toilet and 7-day diarrhea. The following variables were removed: sesind wu3rt tf2a tf4 cd3 hw1 and wu4. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(wu1) + \beta_2(tf5) + \beta_3(tf5a) + \beta_4(cd6) + \varepsilon$$

3.2. Shared toilet and 7-day diarrhea fitting WHO case definition. The following variables were removed: sesind wu3rt tf2a tf4 cd3 hw1 and wu4.

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(tf5) + \beta_2(wu1) + \beta_3(tf5a) + \beta_4(cd6) + \varepsilon$$

4.1. Water source and 7-day diarrhea. The following variables were removedsesind tf4 tf5a cd3 hw1 and wu4. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(tf5) + \beta_4(cd6) + \beta_5(wu3rt) + \varepsilon$$

To determine the association between drinking water ladder category (dwladder) and diarrhea, the following equation was used, after dropping sesind tf5 tf4 tf5a wu3rt cd3 hw1 and wu4:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(dwladder) + \beta_2(saladder) + \beta_3(cd6) + \varepsilon$$

4.2. Water source and 7-day diarrhea fitting WHO case definition. The following variables were removed: tf5 tf5a cd3 hw1 and wu4. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(wu1) + \beta_2(tf4) + \beta_3(tf2a) + \beta_4(SES) + \beta_5(cd6) + \beta_6(wu3rt) + \varepsilon$$

To determine the association between drinking water ladder category and diarrhea (adhering to WHO case definition), the following equation was used, after dropping sesind tf5 tf4 tf5a wu3rt cd3 hw1 and wu4:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(dwladder) + \beta_2(saladder) + \beta_3(cd6) + \varepsilon$$

5.1. Round trip time to fetch water and 7-day diarrhea. The following variables were removed: tf5 sesind tf4 tf5a cd3 hw1 and wu4. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(wu3rt) + \beta_2(tf2a) + \beta_3(cd6) + \beta_4(wu1) + \varepsilon$$

5.2. Round trip time to fetch water and 7-day diarrhea fitting WHO case definition. The following variables were removed: tf2a sesind tf4 tf5a cd3 hw1 wu4. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(wu3rt) + \beta_2(tf5) + \beta_3(cd6) + \beta_4(wu1) + \varepsilon$$

6.1. SES Index and 7-day diarrhea. The following variables were removed: tf5 tf4 wu3rt tf5a cd3 and hw1. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(SES) + \beta_4(cd6) + \varepsilon$$

6.2. SES Index and 7-day diarrhea fitting WHO case definition. The following variables were removed: tf5 tf4 wu3rt tf5a cd3 and hw1. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(tf2a) + \beta_2(wu1) + \beta_3(SES) + \beta_4(cd6) + \varepsilon$$

Table 5.1. Summary of household-level sociodemographic, water, sanitation and hygiene factors.

Variable	Intervention Group [n (95%CI)] (n=789)	Control Group [n (95%CI)] (n=793)	Total [n (95%CI)] (n=1582)
Mean (95%CI) number of individuals per household	5.37 (5.09-5.64)	5.40 (5.17-5.63)	5.38 (5.17-5.59)
Mean (95%CI) number of children <5 years of age per household	1.45 (1.36-1.54)	1.46 (1.38-1.53)	1.45 (1.39-1.52)
Number (%, 95%CI) of female heads of household	80 (13.03, 9.35-17.87)	71 (12.55, 8.78-17.61)	151 (12.90, 9.94-16.57)
Socioeconomic Index			
Lowest	56 (16.90, 13.15-21.46)	66 (22.11, 17.04-28.17)	122 (18.21, 15.04-21.88)
Second	61 (19.92, 16.02-24.50)	61 (20.71, 16.12-26.20)	122 (20.12, 16.92-23.76)
Middle	59 (18.11, 14.39-22.53)	59 (19.85, 15.53-25.02)	118 (18.54, 15.50-22.04)
Fourth	69 (21.39, 17.43-25.97)	55 (19.98, 15.60-25.21)	124 (21.04, 17.82-24.66)
Highest	74 (23.67, 18.95-29.15)	47 (17.35, 12.20-24.09)	121 (22.09, 18.24-26.48)
Primary water source			
Piped water into dwelling	4 (0.46, 0.14-1.51)	2 (0.23, 0.06-0.92)	6 (0.40, 0.14-1.11)
Piped water into yard/plot	4 (0.46, 0.14-1.51)	1 (0.11, 0.02-0.82))	5 (0.37, 0.12-1.11)
Public tap/standpipe	137 (16.81, 11.67-23.60)	175 (21.76, 16.26-28.48)	312 (18.14, 13.95-23.25)
Hand pump/borehole	15 (2.24, 0.68-7.17)	15 (1.72, 0.54-5.39)	30 (2.10, 0.80-5.41)
Protected dug well/covered well	9 (1.11, 0.36-3.37)	1 (0.16, 0.02-1.17)	10 (0.86, 0.29-2.47)
Unprotected dug well/covered well	8 (1.36, 0.66-2.78)	10 (1.18, 0.61-2.26)	18 (1.31, 0.74-2.31)
Protected spring	443 (55.96, 48.43-63.23)	392 (49.94, 43.69-56.19)	835 (54.34, 48.60-59.96)
Unprotected spring	126 (16.24, 12.02-21.58))	146 (18.07, 13.64-23.54)	272 (16.73, 13.33-20.80)
Rainwater	1 (0.11, 0.02-0.82)	1 (0.11, 0.02-0.82)	2 (0.11, 0.02-0.53)
Pond/lake	10 (1.20, 0.63-2.30)	10 (1.31, 0.64-2.63)	20 (1.23, 0.74-2.04)
Stream/river	30 (3.73, 2.06-6.66)	38 (5.08, 3.14-8.10)	68 (4.09, 2.67-6.22)
Roundtrip time to fetch water			
0-5 minutes	66 (8.64, 6.51-11.38)	52 (6.82, 4.84-9.53)	118 (8.15, 6.47-10.23)
6-15 minutes	221 (27.62, 23.93-31.65)	189 (24.15, 20.34-28.42)	410 (26.69, 23.77-29.82)
16-30 minutes	303 (39.47, 35.44-43.66)	339 (42.63, 38.74-46.61)	642 (40.33, 37.18-43.55)
More than 30 minutes	190 (24.26, 20.41-28.58)	207 (26.39, 22.62-30.55)	397 (24.84, 21.80-28.14)
Household treats water	156 (20.11, 16.54-24.23)	133 (16.64, 13.48-20.38)	289 (19.18, 16.39-22.31)
Method of water treatment			
Boiling	103 (78.09, 63.66-87.88)	113 (88.05, 82.30-92.12)	216 (80.75, 69.92-88.32)
Chlorination	19 (18.20, 9.04-33.25)	8 (5.64, 2.82-10.95)	27 (14.85, 7.82-26.38)
Let stand and settle	1 (1.00, 0.14-7.03)	1 (1.02, 0.15-6.59)	2 (1.01, 0.21-4.61)
Boiling and chlorination	3 (2.71, 0.57-11.87)	3 (2.14, 0.72-6.17))	6 (2.56, 0.75-8.37)
Boiling and/or filtering	0	3 (2.14, 0.70-6.37)	3 (0.57, 0.18-1.76)
Frequency of treatment			
Everyday	16 (9.80, 4.87-18.74)	11 (8.76, 4.52-16.28)	27 (9.56, 5.43-16.28)
2-6 times per week	90 (57.53, 49.17-65.86)	82 (61.75, 51.24-71.26)	172 (58.67, 51.67-65.34)

Once per week	40 (26.36, 20.05-33.80)	33 (24.66, 17.21-34.01)	73 (25.96, 20.73-31.98)
Rarely (less than once per week)	10 (6.11, 3.13-11.60)	7 (4.83, 2.41-9.47)	17 (5.81, 3.34-9.92)
Drinking water for child is treated	24 (3.31, 2.07-5.27)	24 (3.15, 1.98-4.97)	48 (3.27, 2.26-4.69)
Method of water treatment			
Boiling	20 (82.75, 59.11-94.09)	18 (74.40, 53.39-88.05)	38 (80.54, 63.76-90.68)
Chlorination	3 (11.68, 3.01-36.04)	4 (15.93, 5.15-39.81)	7 (12.81, 4.86-29.70)
Let stand and settle	0	1 (3.98, 0.63-21.35)	1 (1.06, 0.15-7.11)
Boiling and chlorination	1 (5.56, 0.75-31.52)	0	1 (4.09, 0.56-24.43)
Sanitation			
Pit latrine with slab	238 (33.23, 28.93-37.82)	245 (33.90, 29.28-38.86)	483 (33.41, 30.00-37.00)
Pit latrine with no slab/open pit	454 (63.02, 57.78-67.96)	416 (58.80, 53.33-64.06)	870 (61.89, 57.81-65.81)
Ventilated pit latrine	13 (2.26, 1.07-4.73)	15 (1.98, 0.83-4.66)	28 (2.19, 1.20-3.98)
No toilet/bush	5 (0.68, 0.29-1.61)	3 (0.48, 0.15-1.52)	8 (0.63, 0.30-1.29)
Composting toilet	5 (0.68, 0.29-1.61)	30 (4.45, 2.39-8.13)	35 (1.69, 1.02-2.79)
Toilet area cleanliness			
Fecal matter observed around toilet	326 (45.68, 40.96-50.47)	313 (44.72, 39.78-49.76)	639 (45.42, 41.71-49.18)
No fecal matter	384 (54.32, 49.53-59.04)	392 (55.28, 50.24-60.22)	776 (54.58, 50.82-58.29)
Toilet location (observed)			
Inside house	1 (0.14, 0.02-1.01)	0	1 (0.10, 0.01-0.74)
Inside compound	84 (12.06, 9.20-15.65)	104 (15.04, 11.36-19.65)	188 (12.85, 10.48-15.68)
Inside plot	552 (77.08, 72.45-81.13)	512 (71.65, 66.48-76.31)	1064 (75.62, 72.02-78.90)
Outside plot (neighbor)	76 (10.73, 8.22-13.88)	92 (13.31, 10.00-17.50)	168 (11.42, 9.32-13.92)
Shared toilet	205 (27.21, 22.80-32.12)	206 (26.63, 22.15-31.64)	411 (27.05, 23.57-30.84)
Reported handwashing station	19 (2.62, 1.57-4.34)	24 (2.92, 1.92-4.41)	43 (2.70, 1.84-3.94)
Able to show handwashing station (out of those who reported handwashing station)	18 (95.61, 75.18-99.36)	22 (92.11, 72.51-98.10)	40 (94.59, 82.66-98.46)
Water available for handwashing (out of households able to show handwashing station)	10 (58.13, 31.97-80.40)	12 (56.72, 34.24-76.74)	22 (57.73, 37.58-75.60)
No soap present for handwashing	15 (90.41, 67.04-97.76)	18 (78.60, 53.92-92.01)	33 (86.96, 71.04-94.77)

Table 5.2: Respondent-defined period- and point-prevalence of diarrheal disease by age and gender at baseline between treatment arms

Age (months) Gender	7-day period prevalence of diarrhea			Point prevalence of diarrhea		
	Intervention Arm	Control Arm	Both arms	Intervention Arm	Control Arm	Both arms
0-11	46 (20.60, 15.26-27.20)	36 (13.91, 10.07-18.91)	82 (18.72, 14.65-23.60)	18 (7.31, 4.65-11.30)	10 (3.62, 1.93-7.24)	28 (6.27, 4.24-9.17)
Male	23 (21.38, 13.90-31.42)	16 (11.58, 7.10-18.31)	39 (18.47, 12.91-25.72)	9 (7.43, 3.81-13.98)	4 (2.64, 0.96-7.05)	13 (6.00, 3.32-10.61)
Female	23 (19.82, 12.94-29.14)	20 (16.61, 10.89-24.51)	43 (18.97, 13.52-25.97)	9 (7.19, 3.75-13.34)	6 (4.75, 2.11-10.31)	15 (6.54, 3.79-11.05)
12-23	44 (19.87, 14.94-25.95)	42 (17.67, 13.22-23.22)	86 (19.29, 15.39-23.89)	13 (5.27, 3.06-8.94)	11 (4.12, 2.30-7.24)	24 (4.96, 3.19-7.64)
Male	22 (22.37, 15.06-31.88)	18 (15.10, 9.51-23.13)	40 (20.19, 14.64-27.18)	8 (7.87, 3.95-15.07)	3 (1.92, 0.61-5.89)	11 (6.09, 3.24-11.17)
Female	22 (17.96, 11.79-26.39)	24 (20.33, 13.88-28.78)	46 (18.53, 13.40-25.05)	5 (3.27, 1.37-7.59)	8 (6.39, 3.25-12.15)	13 (4.02, 2.21-7.19)
24-59	77 (13.34, 10.66-16.57)	69 (10.80, 8.54-13.58)	146 (12.67, 10.57-15.12)	16 (2.74, 1.65-4.52)	12 (2.05, 1.12-3.71)	28 (2.56, 1.68-3.87)
Male	38 (12.77, 9.24-17.40)	43 (13.62, 10.11-18.10)	81 (12.99, 10.14-16.50)	9 (2.97, 1.52-5.72)	9 (3.10, 1.54-6.15)	18 (3.00, 1.78-5.02)
Female	39 (13.92, 10.16-18.78)	26 (7.96, 5.45-11.48)	65 (12.34, 9.43-15.97)	7 (2.51, 1.16-5.34)	3 (0.99, 0.32-3.05)	10 (2.10, 1.06-4.15)
Total Male	83 (16.48, 13.24-20.33)	77 (13.38, 10.76-16.52)	160 (15.62, 13.12-18.48)	26 (4.87, 3.29-7.15)	16 (2.73, 1.62-4.56)	42 (4.27, 3.06-5.94)
Total Female	84 (16.08, 12.96-19.78)	70 (12.50, 9.92-15.63)	154 (15.15, 12.71-17.97)	21 (3.67, 2.35-5.69)	17 (2.97, 1.85-4.74)	38 (3.49, 2.43-4.99)
Total	168 (16.11, 13.85-18.66)	152 (13.46, 11.51-15.68)	320 (15.39, 13.64-17.33)	48 (4.27, 3.20-5.66)	34 (2.94, 2.06-4.18)	82 (3.91, 3.08-4.96)

Table 5.3a: Multivariate logistic regression model of the measure of association between water, sanitation and hygiene factors with 7-day respondent-defined period prevalence of reported diarrhea in children under-5.

Primary cook reported child diarrhea in last 7 days fitting						
	Variable	**Unweighted cases/population**	**Wt. crude prevalence (%) (95%CI)**	**Crude prevalence ratio (95% CI)**	**Wt. model-adjusted predicted marginal prevalence (%) (95%CI)**	**Adjusted predicted marginal prevalence ratio (95%CI)**
Water source classification	Surface water	25/120	22.60 (15.27-32.12)	1.42 (0.95-2.11)	20.27 (12.84-30.49)	1.25 (0.79-1.98)
	Unimproved	63/371	16.37 (12.26-21.50)	1.03 (0.76-1.39)	15.98 (11.83-21.23)	0.99 (0.72-1.36)
	Improved	240/1564	15.94 (13.87-18.26)	REF	16.19 (13.89-18.78)	REF
Round-trip time to water source	6-15 minutes	81/435	16.20 (12.24-21.12)	1.17 (0.68-2.02)	15.42 (11.44-20.47)	1.33 (0.74-2.38)
	16-30 minutes	132/697	16.96 (13.68-20.83)	1.22 (0.72-2.09)	17.70 (14.19-21.85)	1.53 (0.87-2.67)
	More than 30 minutes	92/457	16.50 (12.96-20.78)	1.19 (0.71-2.00)	19.14 (15.23-23.78)	1.65 (0.94-2.91)
	0-5 minutes	21/127	13.86 (8.58-21.63)	REF	11.59 (6.91-18.80)	REF
Sanitation classification	Open defecation	2/9	-	-	-	-
	Unimproved	166/1126	15.17 (12.73-17.99)	0.93 (0.71-1.22)	15.93 (13.51-18.69)	0.92 (0.70-1.19)
	Shared	46/232	21.29 (15.05-29.22)	1.30 (0.87-1.94)	22.64 (16.48-30.26)	1.30 (0.89-1.90)
	Improved, not shared	85/483	16.36 (13.07-20.28)	REF	17.37 (13.93-21.45)	REF
Toilet area cleanliness	Fecal matter observed around toilet	140/824	17.48 (14.54-20.87)	1.16 (0.90-1.49)	18.69 (15.76-22.03)	1.18 (0.93-1.51)
	No fecal matter	156/1016	15.12 (12.58-18.06)	REF	15.79 (13.19-18.79)	REF
Shared toilet (all toilet types)	Yes	95/518	20.87 (17.33-24.92)	1.41 (1.10-1.82)	23.54 (18.35-29.67)	1.53 (1.12-2.09)
	No	224/1467	14.76 (12.56-17.28)	REF	15.38 (13.09-17.99)	REF
Socioeconomic quintiles	Lowest	72/417	19.14 (14.95-24.17)	1.38 (0.97-1.96)	21.49 (16.59-27.36)	1.58 (1.12-2.22)
	Second	69/417	15.88 (12.10-20.57)	1.15 (0.78-1.68)	17.90 (13.81-22.89)	1.32 (0.90-1.92)
	Middle	74/413	18.58 (14.56-23.41)	1.34 (0.91-1.98)	18.67 (14.66-23.47)	1.37 (0.94-2.00)
	Fourth	59/415	14.19 (10.59-18.76)	1.02 (0.68-1.53)	15.30 (11.42-20.21)	1.12 (0.75-1.68)
	Highest	54/396	13.87 (10.34-18.35)	REF	13.61 (10.17-17.99)	REF

Table 5.3b: Multivariate logistic regression model of the measure of association between water, sanitation and hygiene factors with 7-day WHO-defined period prevalence of reported diarrhea in children under-5.

Primary cook reported child diarrhea in last 7 days fitting WHO case definition						
	Variable	**Unweighted cases/population**	**Wt. crude prevalence (%) (95%CI)**	**Crude prevalence ratio (95% CI)**	**Wt. model-adjusted predicted marginal prevalence (%) (95%CI)**	**Adjusted predicted marginal prevalence ratio (95%CI)**
Water source classification	Surface water	24/123	20.86 (13.79-30.28)	1.42 (0.93-218)	20.53 (13.57-29.83)	1.40 (0.91-2.14)
	Unimproved	50/389	12.99 (9.28-17.91)	0.89 (0.63-1.25)	13.03 (9.22-18.11)	0.89 (0.62-1.27)
	Improved	228/1658	14.65 (12.66-16.89)	REF	14.71 (12.65-17.04)	REF
Round-trip time to water source	6-15 minutes	70/543	13.09 (10.19-16.65)	0.79 (0.48-1.29)	12.68 (9.69-16.42)	0.82 (0.51-1.30)
	16-30 minutes	119/888	14.53 (11.65-17.98)	0.88 (0.55-1.40)	15.45 (12.41-19.07)	0.99 (0.63-1.58)
	More than 30 minutes	87/568	16.51 (12.84-20.97)	1.00 (0.61-1.64)	18.00 (14.14-22.63)	1.16 (0.72-1.87)
	0-5 minutes	27/158	16.56 (10.61-24.94)	REF	15.54 (10.07-23.20)	REF
Sanitation classification	Open defecation	2/10	-	-	-	-
	Unimproved	157/1193	13.82 (11.67-16.28)	0.97 (0.71-1.32)	14.39 (12.19-16.90)	0.90 (0.66-1.24)
	Shared	47/250	19.45 (13.56-27.08)	1.36 (0.87-2.14)	20.33 (13.95-28.66)	1.27 (0.80-2.03)
	Improved, not shared	74/507	14.26 (10.84-18.55)	REF	15.95 (12.13-20.70)	REF
Toilet area cleanliness	Fecal matter observed around toilet	144/891	16.67 (13.68-20.16)	1.31 (0.98-1.76)	18.01 (14.96-21.53)	1.35 (1.02-1.80)
	No fecal matter	132/1058	12.68 (10.36-15.44)	REF	13.34 (10.93-16.18)	REF
Shared toilet (all toilet types)	Yes	83/556	17.11 (13.63-21.26)	1.23 (0.94-1.61)	21.43 (16.50-27.35)	1.54 (1.13-2.10)
	No	213/1543	13.92 (11.88-16.25)	REF	13.96 (11.95-16.24)	REF
Socioeconomic quintiles	Lowest	72/436	20.11 (15.99-24.99)	1.59 (1.13-2.24)	23.06 (18.10-28.89)	1.83 (1.27-2.65)
	Second	61/431	14.06 (10.58-18.45)	1.11 (0.77-1.61)	15.76 (12.03-20.38)	1.25 (0.85-1.84)
	Middle	64/438	14.90 (11.36-19.30)	1.18 (0.82-1.70)	15.58 (11.96-20.04)	1.24 (0.86-1.78)
	Fourth	53/435	11.52 (8.49-15.44)	0.91 (0.63-1.32)	12.41 (9.10-16.71)	0.99 (0.67-1.45)
	Highest	54/436	12.65 (9.65-16.41)	REF	12.58 (9.51-16.47)	REF

Table 5.4: Multivariate logistic regression model of the measure of association between age group and 7-day period prevalence of respondent-defined and WHO-defined diarrhea in children under-5.

	Primary cook reported child diarrhea in last 7 days			Primary cook reported child diarrhea in last 7 days, fitting WHO case definition		
Age Group	Wt. crude prevalence, % (95% CI)	Predicted marginal prevalence, % (95%CI)	Predicted marginal prevalence ratio (95%CI)	Wt. crude prevalence, % (95%CI)	Predicted marginal prevalence, % (95%CI)	Predicted marginal prevalence ratio (95%CI)
0-5 months	15.82 (10.25-23.61)	17.77 (11.44-26.55)	2.53 (1.26-5.07)	12.90 (8.04-20.07)	14.80 (9.23-22.89)	2.19 (1.03-4.65)
6-11 months	24.87 (18.77-32.17)	25.52 (19.31-32.91)	3.63 (2.05-6.41)	22.71 (17.29-29.24)	22.21 (16.81-28.74)	3.28 (1.74-6.17)
12-23 months	20.57 (16.44-25.41)	20.49 (16.56-25.08)	2.91 (1.58-5.37)	18.50 (14.75-22.95)	19.92 (16.04-24.45)	2.94 (1.48-5.83)
24-35 months	15.74 (12.25-19.99)	17.33 (13.48-21.99)	2.46 (1.37-4.44)	13.72 (10.47-17.77)	14.71 (11.21-19.07)	2.17 (1.17-4.03)
36-47 months	12.89 (9.79-16.79)	13.60 (10.38-17.63)	1.93 (1.03-3.62)	12.01 (9.09-15.69)	13.03 (9.87-17.01)	1.92 (0.95-3.88)
48-60 months	6.09 (3.42-10.64)	7.04 (3.94-12.27)	REF	6.38 (3.52-11.29)	6.77 (3.57-12.48)	REF

Table 5.5a: Multivariate logistic regression model of the measure of association between water quality risk category and 7-day period prevalence of respondent-defined diarrhea in children under-5.

Water Quality	Primary cook reported child diarrhea in last 7 days			
	Wt. crude prevalence % (95%CI)	Crude prevalence ratio (95%CI)	Model-adjusted predicted marginal prevalence % (+/- SE)	Adjusted predicted marginal prevalence ratio (95%CI)
Low risk (1-10 CFU/100mL)	15.46 (10.83-21.59)	1.08 (0.68-1.73)	15.98 (11.30-22.11)	1.03 (0.65-1.64)
Intermediate risk (11-100 CFU/100mL)	16.80 (13.08-21.32)	1.18 (0.80-1.74)	18.05 (14.01-22.95)	1.17 (0.79-1.71)
High risk (101-1000 CFU/100mL)	15.51 (12.39-19.24)	1.09 (0.76-1.56)	16.37 (13.09-20.27)	1.06 (0.74-1.51)
WHO Standard (0 CFU/100mL)	14.28 (10.63-18.90)	REF	15.50 (11.66-20.30)	REF

Table 5.5b: Multivariate logistic regression model of the measure of association between water quality risk category and 7-day period prevalence of WHO-defined diarrhea in children under-5, applying the WHO case definition.

Water Quality	Child diarrhea in last 7 days, fitting WHO Case Definition			
	Wt. crude prevalence % (95%CI)	Crude prevalence ratio (95%CI)	Model-adjusted predicted marginal prevalence % (+/- SE)	Adjusted predicted marginal prevalence ratio (95%CI)
Low risk (1-10 CFU/100mL)	13.39 (9.57-18.43)	0.87 (0.57-1.34)	14.15 (10.14-19.40)	0.80 (0.52-1.24)
Intermediate risk (11-100 CFU/100mL)	13.89 (10.80-17.68)	0.90 (0.63-1.30)	13.56 (10.33-17.61)	0.77 (0.53-1.12)
High risk (101-1000 CFU/100mL)	13.25 (10.51-16.59)	0.86 (0.61-1.22)	14.23 (11.13-18.02)	0.80 (0.55-1.17)
WHO Standard (0 CFU/100mL)	15.38 (11.73-19.92)	REF	17.69 (13.42-22.96)	REF

Table 5.6: Multinomial logistic regression model with generalized logit link to assess the relationship between various water, sanitation and demographic factors and water quality.

	In compliance with WHO Standards (<1 CFU/100mL)			High risk (>100 CFU/100mL)		
Variable	Weighted crude prevalence % (95%CI)	Model-adjusted predicted marginal prevalence% (95%CI)	Model-adjusted predicted marginal prevalence ratio (95%CI)	Weighted crude prevalence % (95%CI)	Model-adjusted predicted marginal prevalence% (95%CI)	Model-adjusted predicted marginal prevalence ratio (95%CI)
Water Source*						
Surface water	6.43 (2.38-16.21)	6.63 (1.82-21.34)	0.25 (0.07-0.88)	68.55 (52.62-81.06)	67.40 (49.79-81.17)	2.00 (1.45-2.76)
Unimproved water source	23.46 (14.07-36.44)	17.76 (11.80-25.84)	0.67 (0.43-1.03)	43.95 (32.51-56.07)	47.25 (35.00-59.84)	1.40 (1.00-1.97)
Other improved water source	29.13 (23.69-35.23)	26.68 (20.87-33.43)	REF	31.51 (26.06-37.52)	33.71 (27.17-40.93)	REF
RT time to water source						
6-15 minutes	21.32 (14.51-30.19)	17.83 (12.24-25.22)	0.80 (0.41-1.58)	37.10 (27.68-47.61)	40.00 (30.24-25.22)	0.93 (0.59-1.45)
16-30 minutes	31.78 (24.25-40.40)	31.26 (23.79-39.84)	1.41 (0.75-2.65)	34.28 (27.85-41.34)	35.46 (28.87-39.84)	0.82 (0.54-1.26)
More than 30 minutes	26.25 (18.06-36.48)	25.55 (17.28-36.05)	1.15 (0.59-2.24)	35.77 (26.07-46.81)	36.15 (28.87-47.63)	0.84 (0.52-1.36)
0-5 minutes	22.89 (13.35-36.38)	22.23 (12.08-37.28)	REF	40.26 (25.39-57.16)	43.06 (27.83-59.72)	REF
Toilet type*						
Unimproved facilities	26.98 (20.67-34.38)	27.14 (20.88-34.45)	1.23 (0.83-1.83)	36.32 (30.01-43.14)	36.31 (30.04-43.08)	0.94 (0.68-1.30)
Improved, shared	21.51 (12.97-33.52)	21.03 (12.69-32.77)	0.96 (0.54-1.70)	37.72 (25.08-52.27)	43.24 (30.72-56.68)	1.12 (0.78-1.61)
Improved, not shared	21.01 (15.10-28.44)	22.02 (15.61-30.12)	REF	39.61 (29.95-50.16)	38.59 (28.63-49.61)	REF
Toilet location						
Inside Plot	23.75 (18.56-29.86)	23.09 (17.66-29.60)	0.71 (0.44-1.15)	38.27 (32.16-44.77)	39.43 (32.73-46.57)	1.31 (0.78-2.20)
Outside Plot	30.29 (18.27-45.80)	15.94 (7.85-29.68)	0.49 (0.22-1.12)	32.96 (22.48-45.46)	36.42 (23.69-51.39)	1.21 (0.64-2.28)
Inside Compound	28.14 (17.17-42.53)	32.49 (19.56-48.79)	REF	33.63 (20.46-49.95)	30.12 (17.25-47.14)	REF
Toilet area cleanliness						
Fecal matter observed around toilet	24.00 (18.07-31.13)	22.41 (16.03-30.41)	0.93 (0.61-1.40)	38.35 (30.78-46.52)	38.69 (30.43-47.67)	1.03 (0.77-1.37)
No fecal matter	25.79 (19.38-33.45)	24.23 (17.94-31.87)	REF	35.97 (29.20-43.35)	37.63 (30.10-45.82)	REF
Shared toilet						
Yes	36.32 (26.72-47.16)	29.76 (19.30-42.88)	1.39 (0.86-2.23)	29.50 (22.03-38.27)	37.54 (26.78-49.69)	0.98 (0.71-1.37)
No	23.29 (18.43-28.98)	21.49 (16.38-27.65)	REF	38.00 (32.13-44.25)	38.20 (31.62-45.24)	REF
Household treats water						
Yes	28.93 (18.91-41.54)	20.99+/-3.72	0.98 (0.57-1.68)	40.48 (28.90-53.23)	39.81+/-5.09	1.22 (0.90-1.65)
No	25.92 (20.83-31.76)	26.27+/-2.11	REF	34.20 (29.05-39.77)	34.74+/-2.32	REF
Designated handwashing location						
Yes	40.96 (22.36-62.56)	55.22 (30.14-77.90)	2.42 (1.47-3.98)	22.50 (10.71-41.27)	19.35 (6.96-43.48)	0.51 (0.19-1.33)
No	26.28 (21.47-31.73)	22.81 (18.01-28.44)	REF	35.68 (30.55-41.16)	38.31 (32.23-44.78)	REF
Persons per sleeping room						
3 to 5	25.58 (19.67-32.55)	26.06 (20.14-33.00)	0.94 (.65-1.34)	38.07 (31.27-45.37)	37.84 (31.01-45.15)	1.24 (0.90-1.70)
6 to 8	6.94 (1.52-26.47)	8.34 (1.75-31.66)	0.30 (0.07-1.39)	66.87 (34.59-88.51)	61.82 (30.44-85.70)	2.02 (1.14-3.59)
1 to 2	27.60 (19.51-37.47)	27.86 (20.20-37.09)	REF	27.81 (20.48-36.56)	30.55 (23.02-39.30)	REF
Socioeconomic quintiles						
Lowest	30.47 (20.15-43.21)	20.58 (11.74-33.56)	0.96 (0.52-1.75)	29.75 (22.24-38.55)	36.67 (26.01-48.82)	0.95 (0.63-1.44)
Second	30.68 (21.48-41.72)	28.18 (19.08-39.51)	1.31 (0.84-2.06)	34.50 (24.59-45.96)	36.02 (25.52-48.05)	0.93 (0.61-1.42)
Middle	24.28 (16.63-34.01)	30.40 (20.79-42.08)	1.42 (0.93-2.15)	36.86 (26.17-49.03)	34.63 (23.77-47.38)	0.90 (0.56-1.45)
Fourth	24.89 (16.41-35.87)	19.84 (13.84-27.62)	0.92 (0.58-1.48)	41.80 (32.35-51.89)	46.82 (36.96-56.93)	1.21 (0.83-1.77)
Highest	22.35 (14.97-31.99)	21.47 (14.64-30.36)	REF	35.21 (24.68-47.40)	38.64 (26.99-51.76)	REF

*Piped water on premises and open defecation categories not included due to low sample size.

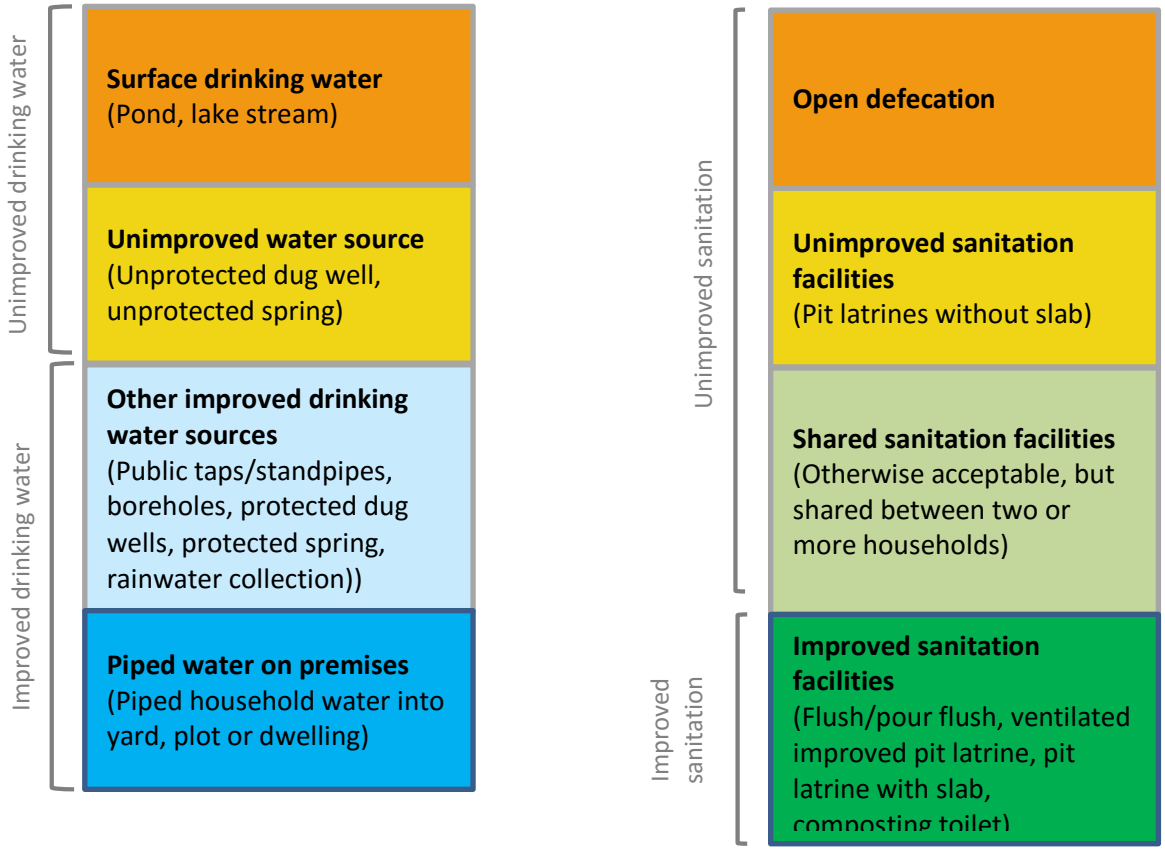


Fig. 5.1. WHO JMP water and sanitation ladders (obtained from figures presented in the WHO JMP 2015 Sanitation and Drinking Water Update [60]

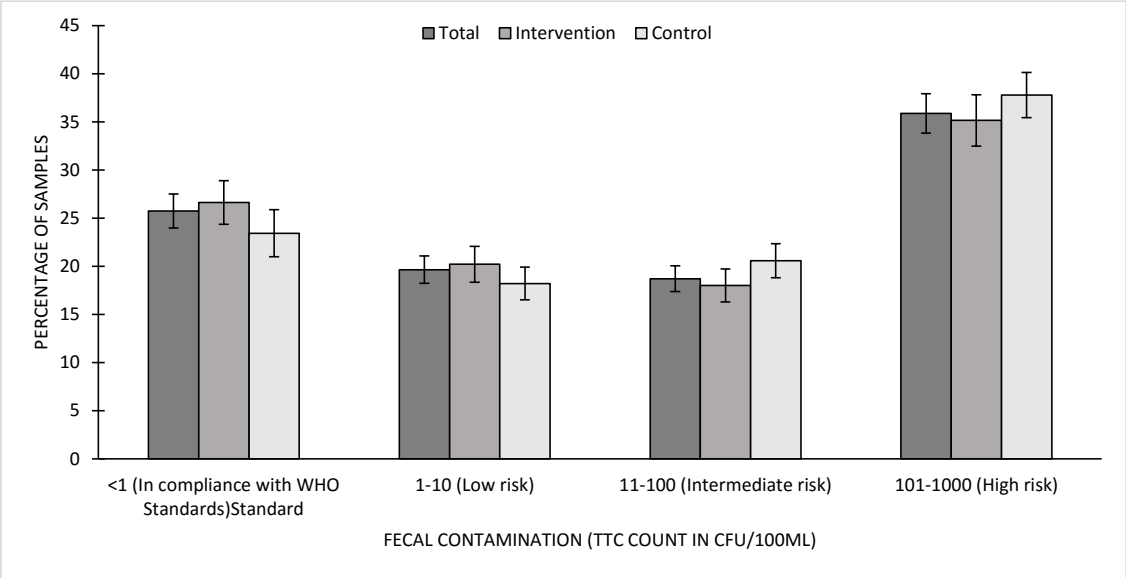


Figure 5.2: Water quality assessments at baseline by treatment assignment

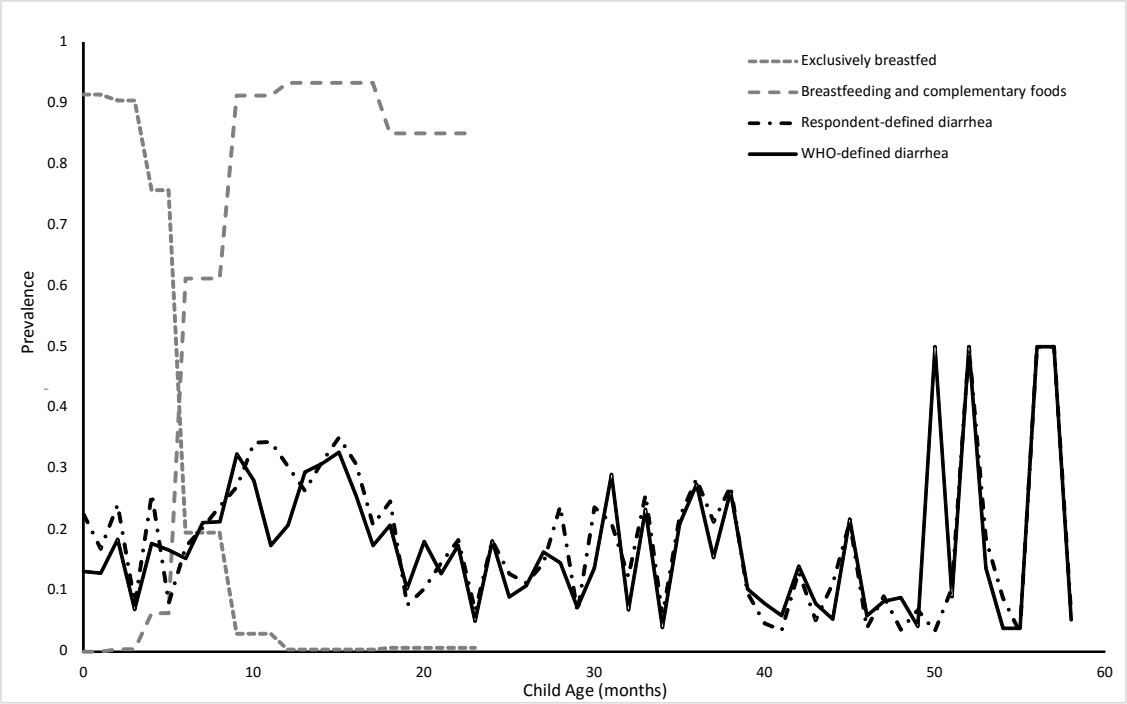


Figure 5.3: Respondent-defined and WHO-defined predicted marginal 7-day diarrhea period prevalence by child age, overlaid by anticipated breastfeeding status as given by the 2010 Rwanda DHS.

Supplemental Table 5.1: Relative proportions of levels of various sociodemographic indicators, prior to aggregation through principal component analysis.

Variable	EZ + LS Group (n=789)	Control Group (n=793)	Total (n=1582)
Education			
Household Head			
No formal education	5 (0.64, 0.27-1.51)	1 (0.12, 0.02-0.86)	6 (0.50, 0.22-1.13)
Completed preschool, no primary	5 (0.61, 0.26-1.45)	4 (0.48, 0.18-1.27)	9 (0.58, 0.28-1.17)
Some primary	368 (47.16, 42.88-51.48)	328 (42.46, 38.26-46.78)	696 (45.91, 42.57-49.29)
Completed primary, no secondary	18 (2.35, 1.50-3.66)	27 (4.11, 2.69-6.22)	45 (2.82, 2.05-3.87)
Some secondary	7 (0.88, 0.42-1.80)	7 (0.84, 0.37-1.92)	14 (0.87, 0.49-1.54)
Completed secondary	3 (0.37, 0.12-1.14)	4 (0.48, 0.15-1.59)	7 (0.40, 0.17-0.94)
Missing	359 (48.00, 43.74-52.29)	383 (51.50, 47.38-55.60)	742 (48.93, 45.61-52.26)
Primary Cook			
No formal education	3 (0.37, 0.08-1.63)	3 (0.34, 0.08-1.48)	6 (0.36, 0.11-1.16)
Completed preschool, no primary	5 (0.86, 0.31-2.36)	0	5 (0.63, 0.22-1.73)
Some primary	526 (65.66, 61.29-69.77)	530 (66.17, 62.32-69.81)	1056 (65.79, 62.46-68.98)
Completed primary, no secondary	13 (1.68, 0.99-2.82)	18 (2.90, 1.68-4.94)	31 (2.01, 1.37-2.94)
Some secondary	31 (3.96, 2.81-5.57)	24 (3.11, 1.98-4.87)	55 (3.73, 2.81-4.95)
Completed secondary	4 (0.50, 0.19-1.32)	4 (0.59, 0.21-1.63)	8 (0.52, 0.25-1.10)
Missing	207 (26.98, 22.78-31.62)	214 (26.89, 23.40-30.69)	421 (26.95, 23.71-30.46)
Wealth classification			
Ubudehe 1	120 (15.42, 12.21-19.30)	46 (5.73, 4.00-8.14))	166 (12.83, 10.41-15.73)
Ubudehe 2	632 (81.19, 76.74-84.95)	722 (92.88, 89.99-94.98)	1354 (84.31, 80.98-87.15)
Ubudehe 3 and above	21 (3.00, 1.32-6.84)	7 (1.39, 0.49-3.86)	28 (2.56, 1.23-5.46)
Intervention-eligible, Ubudehe level	3 (0.39, 0.08-1.99)	0	3 (0.29, 0.06-1.45)
Household possessions			
Own plot of land-not for farming	725 (91.68, 88.63-93.97)	727 (91.41, 88.88-93.40)	1452 (91.61, 89.35-93.42)
Own agricultural land	454 (56.49, 50.64-62.16)	448 (56.92, 51.39-62.29)	902 (56.60, 52.08-61.02)
Own house	711 (90.03, 86.87-92.50)	707 (88.94, 85.88-91.41)	1418 (89.74, 87.36-91.72)
Electricity	48 (5.80, 3.71-8.96)	48 (5.92, 3.84-9.01)	96 (5.83, 4.13-8.18)
Radio	264 (32.90, 29.64-36.34)	236 (29.91, 26.38-33.71)	500 (32.10, 29.51-33.71)
Television	10 (1.36, 0.59-3.12)	5 (0.56, 0.24-1.33)	15 (1.15, 0.55-2.38)
Mobile telephone	338 (42.82, 38.75-47.00)	311 (39.38, 35.24-43.67)	649 (41.90, 38.71-45.15)
Mattress	250 (31.68, 27.61-36.05)	221 (28.06, 23.43-33.20)	471 (30.70, 27.45-34.16)
Bicycle	18 (2.08, 1.18-3.64)	15 (1.88, 1.16-3.04)	33 (2.03, 1.30-3.14)
Motorcycle	0	1 (0.11, 0.02-0.82)	1 (0.03, 0.00-0.22)
Boat	2 (0.23, 0.06-0.92)	2 (0.23, 0.06-0.92)	4 (0.23, 0.08-0.68)
Livestock			
Household tends cows	246 (31.02, 26.65-35.75)	246 (31.19, 25.24-35.63)	492 (30.79, 27.30-34.52)
Owns other livestock/poultry	402 (50.21, 45-11-55.30)	434 (54.80, 49.62-59.87)	836 (51.45, 47.46-55.41)
Building/construction of home			
Type of flooring materials			
Earth/sand	757 (95.97, 93.92-97.34)	757 (95.51, 92.78-97.24)	1514 (95.84, 94.26-97.01)
Animal dung	0	1 (0.11, 0.02-0.82)	1 (0.03, 0.00-0.22)
Cement	21 (2.70, 1.72-4.22))	29 (3.63, 2.09-6.24)	50 (2.95, 2.07-4.19)
Bricks	9 (1.09, 0.54-2.17)	4 (0.51, 0.19-1.36)	13 (0.93, 0.51-1.71)
Wall materials			
Wood planks only	1 (0.11, 0.02-0.82))	3 (0.34, 0.11-1.06))	4 (0.18, 0.06-0.53)
Wood planks and mud	202 (25.29, 18.49-33.57)	204 (26.18,19.06-34.81)	406 (25.53, 20.06-31.89)
Mud bricks, not covered with mud	122 (15.46, 12.29-19.27)	106 (13.94, 10.80-17.81)	228 (15.05, 12.54-17.97)
Mud bricks covered with mud	373 (47.03, 40.71-53.44))	403 (50.12, 43.22-57.02)	776 (47.86, 42.87-52.90)
Mud bricks covered with cement	69 (8.99, 6.52-12.26)	58 (7.02, 4.87-10.02)	127 (8.46, 6.51-10.91)
Real/clay bricks, not covered	1 (0.11, 0.02-0.82)	1 (0.11, 0.02-0.82)	2 (0.11, 0.02-0.53)
Real bricks, covered with cement	3 (0.34, 0.11-1.06)	3 (0.36, 0.12-1.10)	6 (0.35, 0.15-0.83)
Wood covered with cement	3 (0.46, 0.14-1.51)	6 (0.79, 0.29-2.10)	9 (0.55, 0.24-1.25)
Roof materials			
Metal sheets	313 (39.14, 30.76-48.22)	272 (35.68, 27.24-45.11)	585 (38.21, 31.60-45.29)
Clay tiles	471 (60.28, 51.32-68.61)	517 (63.98, 54.60-72.39)	988 (61.28, 54.27-67.84)
Plastic sheets	3 (0.36, 0.12-1.10)	2 (0.23, 0.06-0.92)	5 (0.32, 0.12-0.83)

Supplemental Table 5.2: Factor loading scores associated with each variable consolidated in the principal components analysis.

Variable	Coded as:	Factor score
Primary cook education (pced)	0=No formal education or Don't Know or Refused or Missing 1=Nursery 2=Some primary 3=Completed primary 4=Some secondary 5=Completed secondary	0.47078
Household head education (hhed)	0=No formal education or Don't Know or Refused or Missing 1=Nursery 2=Some primary 3=Completed primary 4=Some secondary 5=Completed secondary	0.46429
Electricity (sd4)	0=No 1=Yes	0.68138
Radio (sd5a)	0=No 1=Yes	0.57888
Mobile telephone (sd7a)	0=No 1=Yes	0.73996
Mattress (sd8)	0=No 1=Yes	0.77252
Owns agricultural land (sd13)	0=No 1=Yes	0.16014
Owns home (sd14)	0=No 1=Yes	0.05156
Tends cows (sd15)	0=No 1=Yes	0.19110
Roof type (roof)	0=Plastic sheets or Other 1=Metal sheets 2=Clay tiles	-0.14222
Floor type (floor)	0=Mud/earth or Animal Dung or Other 1=Bricks 2=Cement	0.76075
Wall type (wall)	0=Only wood planks or Other 1=Wood planks and mud 2=Mud bricks-not covered 3=Mud bricks- covered with mud 4=Mud bricks-covered with cement 5=Real bricks-not covered 6=Real bricks-covered with cement 7=Wood covered with cement	0.38565

Supplemental Table 5.3a: Association between 7-day respondent-defined diarrhea prevalence with disaggregated water source and sanitation indicators, reflecting risk attributable to each water source and toilet type.

Primary cook reported child diarrhea in last 7 days						
	Variable	**Unweighted cases/population**	**Wt. crude prevalence (%) (95%CI)**	**Crude prevalence ratio (95% CI)**	**Wt. model-adjusted predicted marginal prevalence (%) (95%CI)**	**Adjusted predicted marginal prevalence ratio (95%CI)**
Water source	Piped water into dwelling	0/8	-	-	-	-
	Piped water into yard/plot	0/6	-	-	-	-
	Public tap/standpipe	72/439	21.39 (16.85-26.77)	1.50 (1.13-1.99)	24.37 (19.20-30.41)	1.74 (1.30-2.33)
	Hand pump/borehole	12/40	22.15 (11.46-38.48)	1.56 (0.82-2.94)	19.58 (9.77-35.38)	1.40 (0.70-2.77)
	Protected dug well/covered well	0/14	-	-	-	-
	Unprotected dug well/covered well	3/23	-	-	-	-
	Unprotected spring	47/366	16.84 (12.48-22.33)	1.18 (0.86-1.63)	15.71 (11.57-20.99)	1.12 (0.80-1.57)
	Rainwater	1/3	-	-	-	-
	Pond/lake	8/32	35.84 (19.55-56.22)	2.52 (1.44-4.41)	27.55 (13.70-47.68)	1.96 (1.02-3.78)
	Stream/river	16/91	18.17 (10.20-30.26)	1.28 (0.72-2.27)	16.21 (8.89-27.72)	1.16 (0.63-2.11)
	Protected spring	143/1148	14.24 (12.06-16.73)	REF	15.32+/-1.38	REF
Toilet type	Pit latrine with slab	102/665	17.69 (14.45-21.46)	1.17 (0.89-1.52)	18.63 (15.32-22.46)	1.19 (0.92-1.53)
	Ventilated pit latrine	7/42	11.85 (4.10-29.73)	0.78 (0.28-2.16)	9.95 (3.20-26.96)	0.64 (0.21-1.89)
	No toilet/bush	2/10	-	-	-	-
	Composting toilet	12/50	32.56 (18.31-50.99)	2.15 (1.26-3.66)	28.31 (16.33-44.41)	1.81 (1.08-3.04)
	Pit latrine with no slab/open pit	157/1193	15.17 (12.73-17.99)	REF	15.67 (13.19-18.51)	REF

Supplemental Table 5.3b: Association between 7-day WHO-defined diarrhea prevalence with disaggregated water source and sanitation indicators, reflecting risk attributable to each water source and toilet type.

Primary cook reported child diarrhea in last 7 days fitting WHO case definition						
	Variable	**Unweighted cases/population**	**Wt. crude prevalence (%) (95%CI)**	**Crude prevalence ratio (95% CI)**	**Wt. model-adjusted predicted marginal prevalence (%) (95%CI)**	**Adjusted predicted marginal prevalence ratio (95%CI)**
Water source	Piped water into dwelling	0/7	-	-	-	-
	Piped water into yard/plot	0/4	-	-	-	-
	Public tap/standpipe	79/343	17.91 (13.69-23.07)	1.33 (0.97-1.83)	20.74 (15.99-26.46)	1.62 (1.18-2.22)
	Hand pump/borehole	12/28	29.96 (23.98-36.72)	2.23 (1.68-2.97)	28.11 (20.73-36.90)	2.20 (1.52-3.18)
	Protected dug well/covered well	0/12	-	-	-	-
	Unprotected dug well/covered well	4/18	-	-	-	-
	Unprotected spring	59/290	13.35 (9.40-18.63)	1.00 (0.69-1.43)	12.97 (9.14-18.08)	1.01 (0.71-1.45)
	Rainwater	1/2	-	-	-	-
	Pond/lake	8/22	32.31 (18.37-50.31)	2.41 (1.40-4.13)	26.86 (15.55-42.28)	2.10 (1.22-3.60)
	Stream/river	17/73	16.69 (9.37-27.97)	1.24 (0.69-2.23)	16.86 (9.72-27.65)	1.32 (0.75-2.31)
	Protected spring	148/1076	13.42 (11.23-15.98)	REF	12.81 (10.70-15.26)	REF
Toilet type	No toilet/bush	2/10	-	-	-	-
	Pit latrine with slab, shared	28/158	18.43 (11.76-27.68)	1.33 (0.83-2.16)	20.61 (13.30-30.53)	1.46 (0.92-2.32)
	Ventilated pit latrine, shared	1/11	-	-	-	-
	Composting toilet, shared	1/15	-	-	-	-
	Pit latrine with slab, not shared	74/507	14.26 (10.84-18.55)	1.03 (0.76-1.41)	15.61 (11.85-20.29)	1.10 (0.81-1.51)
	Ventilated pit latrine, not shared	6/31	13.52 (4.74-32.95)	0.98 (0.36-2.67)	9.07 (1.84-34.70)	0.64 (0.14-3.01)
	Composting toilet, not shared	11/35	37.94 (19.70-60.36)	2.75 (1.56-4.84)	35.17 (19.83-54.34)	2.49 (1.48-4.17)
	Pit latrine with no slab/open pit	157/1193	13.82 (11.67-16.28)	REF	14.15 (11.98-16.65)	REF

Chapter 6 - Predictors of acute respiratory infection, pneumonia and household air pollution; a cross-sectional study in Western Province, Rwanda

Abstract

Acute lower respiratory infection (ALRI) is the largest contributor to childhood mortality in Sub-Saharan Africa. Much of this excess mortality among children under 5 years-old is attributed to household air pollution caused by burning of biomass fuels for cooking and heating. This study sought to assess the association between various demographic and household characteristics and cooking behaviors with ALRI and personal household air pollution exposure, measured by 24-hour mean fine particulate matter ($PM_{2.5}$) concentration. The link between $PM_{2.5}$ exposure and ALRI in this population was also explored along with other potential ALRI comorbidities, including diarrhea and malnutrition. We obtained household-level data from primary cooks including relevant demographic factors and exposures. Child health status was assessed through seven-day recall of specific disease outcomes, and field-based methods were used to identify current cases of ALRI. Age-specific ALRI was highest among children 6-11 months-old compared to children over 48 months-old (RR=4.78, 95%CI: 1.72-13.28). Across all age groups, females had a higher risk for ALRI than males (RR=1.46, 95%CI: 1.03-2.06). A subset of the survey population also participated in a 48-hour personal HAP exposure study. While stove type appeared to be associated with upper respiratory infection, only socioeconomic status appeared to be associated with ALRI, with all four lower quintiles yielding a higher risk of ALRI than the highest SES quintile. Stove type, cooking area, number of meals cooked per day, fuel type and kerosene usage differentially affected personal 24-hour mean $PM_{2.5}$ concentration exposure among primary cooks and children.

An interquartile increase in logarithmically-transformed 24-hour PM2.5 concentration yielded no significant increase in upper respiratory infections but did significantly increase the risk of ALRI in young children (RR=2.99, 95%CI: 1.00-9.25). Adjusted bivariate probit analysis yielded a significant association between diarrhea and ALRI (ρ=0.1785, p=0.004) and an apparent increased risk for ALRI among malnourished children (ρ=-0.1192, p=0.084). Together, this research presents the associations between cooking factors and behaviors associated with $PM_{2.5}$ exposure and the association of ALRI with both $PM_{2.5}$ exposure and other comorbidities.

Introduction

Acute lower respiratory infection (ALRI) is the leading cause of morbidity and mortality among children under-5 years old in Sub-Saharan Africa, outnumbering the total number of childhood deaths attributed to HIV and tuberculosis combined [62]. Despite progress in decreasing global ALRI-attributable morbidity between 1990 and 2013 (Vos, Barber, Bell, et al., 2015), ALRI is still responsible for 4.6% of global disease burden [1] and 4.3 million premature deaths each year [87]. Despite the disproportionate burden of ALRI borne by children throughout Sub-Saharan Africa [86], Rwanda has substantially reduced the under-5 mortality rate between 2000-2015 compared to the 1990's, making it an attractive country for further investment in interventions addressing child health and mortality [88].

Globally, 41% of households representing 2.8 billion people rely mainly on solid fuels for cooking [32]. A systematic review and meta-analysis performed by Dherani et al. (2008) revealed that the odds of pneumonia in children under-5 was 78% higher in

children living in households with high household air pollution (HAP) exposure households vs. children living in low HAP exposure households; a similar association was described by Torres-Dosal et al. (2008), who further linked HAP exposure to tuberculosis and low birth weight. Household air pollution (HAP) due to inefficient biomass-burning cookstoves alone was associated with 81.1 million global disability-adjusted life-years (DALYs) and 2.1 million deaths in 2013 [89], making it the 7th leading risk factor for global burden of disease and the 4th leading risk factor for among all causes for disease burden in Rwanda, after child undernutrition, unsafe sex and unsafe water [89]. Among the top non-communicable contributors to global DALYs, HAP is the only exposure that disproportionately affects children [2]. HAP is the primary contributor of respiratory disease in children under 5 and a leading cause of chronic bronchitis and chronic obstructive pulmonary disease (COPD) among women [2,3]. Further demonstrating this, Further exacerbating the health risks attributed to HAP exposure, HAP has been associated with inflammatory processes linked to cardiovascular sequelae, atherosclerosis and hypertension [43]. Given its enormous impact on child health, HAP is thought to be one of the largest preventable causes of morbidity and mortality in the world [87].

In low- and middle-income countries (LMICs), HAP is predominantly produced from burning solid biomass fuels. The majority of this activity is driven by cooking, although lighting and heating can also result in the production of black carbon smoke [90]. While WHO guidelines recommend that household levels of HAP based on particulate concentrations not exceed 10ug/m^3 for PM2.5 (35 ug/m^3 for the current interim target) and 50ug/m^3 for PM10, in households that burn biomass for cooking [91], concentrations

can exceed 10,000 ug/m^3 during cooking or other periods [3]. In most settings, women undertake the vast majority of cooking responsibilities, and children are often carried on their backs or put to sleep nearby. As a result, both women and children are disproportionately exposed to several periods of intense cooking smoke exposure each day [90].

In addition to the substantial health risks that HAP presents within the home, the environmental impact of household biomass burning has become a target for global carbon emissions reductions. Forty percent of the global population burns biomass fuels in the household [92], and that proportion increases to 90% in rural regions of Sub-Saharan Africa [31]. Recognizing their potential to reduce the contribution of biomass fuel burning to global carbon emissions, improved cookstove interventions have become an attractive mechanism to couple sustainable development with the potential for carbon credit revenue, incentivized through the UN Clean Development Mechanism [92].

However, previous evidence linking improved cookstove distribution to direct health outcomes has been scant [7]. Demonstrable reductions in ARI and ALRI prevalence typically are not observed above 24-hour mean $PM_{2.5}$ concentrations of 200 ug/m^3 in exposure-response curves (Ezzati & Kammen, 2002); therefore, large exposure reductions are needed in affected households to have a presumed health impact [32,94]. Despite the fact that Rwanda is driving the most pronounced reduction of childhood mortality in the world, smoke from biomass-burning stoves contributes substantially to ALRI-associated mortality in Rwanda and in the region as a whole [89]. These HAP

exposures have been targeted as a means to reduce ALRI burden among children throughout the country.

To address both the environmental and disease burden resulting from biomass burning in Rwanda, DelAgua Health Rwanda (DelAgua), under the authority of the Ministry of Health (MOH), Rwanda, implemented a pilot program in 2012 to provide high efficiency biomass cookstoves to approximately 2200 householder in 15 villages in rural Rwanda. The intervention also included advanced water filters that aimed at a substantial disease burden associated with unsafe drinking water. [5]. An evaluation of the pilot project used a parallel household-randomized controlled trial design of three rural villages in order to assess uptake and use of the intervention and the impact of the intervention on air quality near household cooking areas [6]. Two-thirds (66.7%) of intervention households identified the intervention stove as their main cooking stove, but only 23.3% of intervention households reported that their main cooking area was outdoors. Overall, the stoves were associated with a 48% reduction of 24-h PM2.5 concentrations in the cooking area (267 $\mu g/m^3$ vs. 509 $\mu g/m^3$, $p = 0.005$). The reduction was 37% for those cooking indoors ($p=0.08$) and 73% for those cooking outdoors ($p<0.001$) (Rosa 2014).

Based on the initial results, DelAgua and the Rwanda MOH elected to scale up the intervention to cover the poorest third of the population (*Ubudehe* 1 and 2) throughout Western Province (Phase 2). The implementation plan called for delivery to 72 of the 96 sectors (groups of villages that also correspond with catchment areas for primary care clinics), with the balance to be covered approximately one year later. As the MOH and DelAgua agreed to select the initial round randomly to ensure equity, we took advantage of this natural experiment to conduct a sector-level cluster-randomized controlled trial to

assess the impact of the intervention on health outcomes using records maintained by the clinics and CHWs (the "clinic-level RCT"). At the same time, we randomly selected 87 villages from each arm of the sector-level RCT for a nested village-level RCT where we could assess coverage, uptake (use), exposure and other measures of health outcomes (reported, CHW recorded, instrumented and potential blood-based biomarkers) (the "village-level RCT"). The main objective of the village-level RCT is to assess the impact of the intervention and HAP and fecal contamination of drinking water—the main exposures that the intervention aims to mitigate. A sub-study is also designed to investigate possible biomarkers of enteric and respiratory exposures and disease in an effort to develop more objective criteria for assessing these health disorders and the interventions designed to prevent them.

For the village-level RCT, we enrolled 1582 households with children <5 from 174 study villages, evenly distributed between intervention and control arms. We undertook a comprehensive baseline survey that included extensive information from study participants on demographics, water sources and management practices, cooking fuels and cooking practices.

This paper summarizes the results of the baseline study. In addition to exploring balance between the intervention and study arms and describing the study population however, we will explore associations between household and environmental factors and both ARI and ALRI and HAP exposure and between HAP exposure and ALRI. We will also explore co-occurrence of diarrhea, malnutrition and ALRI.

Methods

Setting

Western Province is located in western Rwanda, bordered by the Democratic Republic of the Congo and Lake Kivu to the west, Burundi to the south, Southern Province (Rwanda) to the east/southeast and Northern Province (Rwanda) to the east/northeast. The province is divided into 7 districts: Karongi, Rutsiro, Rubavu, Nyabihu, Ngororero, Rusizi and Nyamasheke. These 7 districts are further divided into 96 administrative sectors, which in turn are divided into 538 cells and 3612 villages. Western Province has the highest percentage of men (17%) and women (26%) who have never obtained any formal education; however, this figure is largely influenced by age. Across the country, for example, 79% of women age 65 and over have no education, compared to 2% of girls between the ages of 10 and 14. We anticipate the same trend in our study population: the older our respondents are, the less likely they are to have had a formal education.

Cooking habits and fuel sources. According to the 2010 Demographic and Health Survey (DHS), 77% of households use wood as cooking fuel country-wide, with the 2nd most common cooking fuel being straw, shrubs or grass (12%); however, the proportion of households that use wood for cooking fuel increases to 83.3% in rural areas, with 12.4% using straw, shrubs or grass and 3.0% using charcoal [12]. Households that use wood fuel for cooking spend, on average, 50 minutes per day collecting wood. Most households in Western Province report that they prefer wood to other sources because it is relatively available (72%) and because it can be obtained without purchasing (17%) [33]. The predominant difficulties reported by residents in Western Province with gathering wood are that the activity is difficult in the rainy season (24%) and that cutting

wood is difficult (21%). Other reported difficulties included the distance to retrieve firewood, the weight during transport and the inaccessibility of the wood [33]. In rural areas, 28.9% of cooks report cooking inside the house, 52.2% report cooking in a separate building, and 18.0% report cooking outside [12].

Selection of Study Participants. The Rwandan MOH and DelAgua developed a plan to roll out the intervention to all *Ubudehe* 1 and 2 households in Western Province; however, due to funding and logistical constraints, the roll out for the beginning of Phase 2 was only feasible for 100,000 households. Our study team opted to take advantage of this limitation to design a rigorous cluster-randomized controlled trial, in which, sectors were chosen as the unit of randomization since they comprise the catchment area for health clinics. The research team randomized the intervention at a 3:1 ratio to 72 sectors,consisting of the 100,000 households the implementer sought to reach at the start of this phase); the control arm, consisting of 24 sectors (about 40,000 households) will receive the intervention after the end of the follow-up period of our study. In this main "sector-level" study, the impact of the intervention will be assessed using sector-level clinic data and community health worker (CHW) records. Nested within this overall sector-level study is a more in-depth assessment of 174 villages selected randomly using population proportional selection (the "village-level sub-study") split between intervention in control arms. This involves air and water sampling together with more extensive exposure and health data collected from up to 10 households in each village. Following enrollment and a baseline assessment from August through December 2014, each participating household and will undergo follow-up visits on a quarterly basis through March 2016.

Within the 87 intervention and 87 control villages selected into the study, the sampling frame was the village-specific intervention distribution list provided by the *Tubeho Neza* program. This list was derived from the Rwandan government's 2012 *Ubudehe* List, which the country uses to identify households eligible for various government-sponsored welfare programs. Households were considered eligible for the intervention if they fell into the lowest socioeconomic tertile, designated as *Ubudehe* levels 1 and 2. In addition to this socioeconomic requirement, households had to have at least one child under 5 years-old and have a primary cook 16 years-old or older to be eligible for selection into our study. Households in highly urbanized areas, such as Gisenyi in Rubavu and Kamembe in Rusizi, were excluded, as their use of coal over wood as cooking fuel precluded them from receiving the wood-burning EcoZoom cookstove.

Within each of the 174 study villages, eligible households assigned a random number from a random number generator and were visited in order of the random number. All households we visited were invited to participate in the baseline study and up to four follow-up rounds of the study for up to 16 months. Eligible households were recruited from each village until either 10 households were enrolled or the eligible population was exhausted. Informed consent for the baseline study was obtained from the primary cook of the household on behalf of all household members, and the primary cooked served as the principal respondent for the survey and all subsequent visits throughout follow-up. If a new primary cook was identified during subsequent visits, the consent procedure was repeated. Households were not provided with any incentive to participate, and their receipt of the *Tubeho Neza* intervention was not contingent upon their participation.

Population stratification and hierarchical selection. Data were collected in a hierarchical manner. The intervention was ultimately randomized by sector, and data were collected within multiple intermediate levels of observation:

Cluster: Sectors, which were randomized at a 3:1 ratio. Sectors are government administration areas; there are a total of 96 sectors in Western Province that lie within 7 districts. Cluster-level effects are mostly relevant for follow-up studies; cluster assignment is accounted for in this cross-sectional baseline assessment through the sample weights assigned to households and individual children and through the use of robust standard errors to account for intracluster-correlation at the village level.

Village: There are 174 villages selected into this study, randomly selected using population proportional selection at a 1:1 intervention-to-control ratio. Village size will be considered as it may contribute to transmission of respiratory and diarrheal disease.

Household: 1582 households were ultimately enrolled across our study, for an average village enrollment size of 9.15 households per village. Potentially relevant factors associated with the outcomes of interest for our analysis at this level include indicators of socioeconomic status and water, sanitation and hygiene factors.

Individual: 2179 children were enrolled across our study, for an average of 1.38 children under 5 years-old enrolled per household. Children over 4 years-old at baseline were only enrolled if a younger child also lived in the house to avoid having whole households age

out of the cohort throughout the anticipated 12 month follow-up period. In general, this resulted in fewer children between 48 to 60 months-old in our study compared to other age groups. Outcome-level data will be collected at this level, in addition to potential confounders such as age, gender and immunization status.

As ensuing follow-up analyses will account for treatment-level effects; clustering was accounted for both at the level of intervention assignment and at the village level. For the purposes of variance estimation, our sample is separated into two strata (intervention and control), with 87 village clusters serving as primary sampling units within each stratum. Sampling within each PSU is assumed to be with replacement, and each household serves as a unit of analysis clustered within each PSU. SAS-Callable SUDAAN (Research Triangle Institute, Research Triangle Park, NC) was used for all logistic and linear regression models, which included weights for each child and incorporated Taylor series linearization and Zeger robust standard errors to account for between-cluster variance.

Data collection. With the exception of households receiving more extensive personal HAP monitoring, participating households of the overall study were visited once during baseline. Data on household demographics, primary cook and household head education, stove type, number of stoves, cooking frequency, cooking area characteristics and other potential covariates were collected during the baseline survey. All survey instruments were first written in English and then underwent a double forward- and backward-translation process to obtain our final survey instruments in Kinyarwanda. All surveys were piloted before use in our study, and pilot participants and survey enumerators were asked to provide their feedback on the comprehensibility of the questions asked. Survey

items were a combination of enumerator observations and participant questions; they were predominantly binary or multiple-choice, although some open-ended questions were asked where appropriate.

During the household visits, enumerators asked primary cooks about current and 7-day binary recall of cough and difficulty breathing for themselves and for each child under 5 years-old in the household. In addition, respondents were asked to provide information on duration of each reported illness (in days; care-seeking behavior for each illness from a CHW or health center; and vaccines received, as confirmed from a vaccination card administered by the Ministry of Health. If current cough in any child under 5 years-old was reported, the child was assessed for 1) chest indrawing, 2) rapid breathing and 3) stridor using the WHO's Integrated Management of Childhood Illness (IMCI) criteria [69]. Rapid breathing was defined as ≥40 breaths per minute in children between 12 and 60 months old and ≥50 breaths per minute in children 2 to 12 months-old. Enumerators identified potential ALRI cases among all enrolled children older than 2 months-old during household visits based on a combination of these lower respiratory sequelae [69] (Figure 6.1), using the IMCI manual used to train community health workers in Rwanda by the MOH. All children with IMCI-classified general danger signs were referred into care through one of the village's designated CHWs. Regardless of the presence of diarrheal disease, middle-upper arm circumference (MUAC) measurements were obtained to identify cases of child malnutrition. Enumerators were trained on the IMCI protocol specific to the Rwandan Ministry of Health by the Ministry's own CHW trainers and pediatric clinic staff.

Personal HAP exposure monitoring.

$PM_{2.5}$ is a well-established marker of personal health risks from combustion-related emissions and has been linked to excess morbidity and mortality globally [32,89]. Carbon monoxide (CO) is also a byproduct of the combustion process, and measurements such as carboxyhemoglobinn (COHb) and exhaled CO concentrations can be used as a proxy for personal exposure to HAP [32,91]. This section details the methods used to ascertain personal HAP exposure in our study population.

Filter processing and weighing. This filter weighing protocol followed standard methods for particulate matter (PM) gravimetric analysis [95]. Prior to all field activities, one 37mm, 2uM polytetrafluoroethylene (PTFE) membrane filter with support ring (Pall Life Sciences, Port Washington, NY) was measured on a microbalance (Cole-Palmer) in a hood that controlled for humidity (30-40%) and temperature (20-23°C) and limited air flow. Before weighing, all filters were conditioned by placing them in their respective petri dishes in a dessicator (BelArt Products, Wayne, NJ) with lithium chloride dessicant for a minimum of 20 hours. After filter conditioning, filters were removed from the dessicator and passed across an electrostatic bar immediately before measurement. All filter measurements were stabilized on the microbalance for a minimum of 15 seconds, and all measurements were performed twice. In the event that two measurements differed by more than 5ug, a third measurement was taken. The mean of all measurements for any given filter was used for this analysis. After the filters were deployed in the field, they were returned to the same laboratory for a post-measurement following identical procedures. $PM_{2.5}$ mass (ug) was calculated by subtracting the pre-weighted mass from the post-weighted mass.

HAP participant selection and set-up. The first two households that appeared on the randomly ordered household list for each village were selected for 24-hour personal HAP exposure monitoring. Within each selected household, one child between 1.5 and 4 years-old and one main cook was selected for this baseline assessment. This age range for children was selected since children under 1.5 years-old are typically too small to carry the 1.4 kg backpack containing HAP equipment, and children under-4 at baseline will not age out of the under-5 cohort over the one year follow-up period. Gravimetric $PM_{2.5}$ exposures were obtained cumulatively over 48 hours through a wearable pump and filter system. The pumps were set to operate one minute on, one minute off, for 48 hours to acquire a cumulative 24-hour mean $PM_{2.5}$ concentration. Our personal HAP monitoring system consists of a programmable TuffPro™ air pump (Casella, Inc., Buffalo, NY), a light sensor logger (HoboWare, Inc.) to help assess compliance with wearing the HAP system, a Harvard Personal Environmental Monitor (HPEM) $PM_{2.5}$ cyclone (Harvard School of Public Health, Boston, MA) designed for an optimal flow rate of 1.8 L/min, the pre-weighed PFTE filter and rubber tubing connecting the HPEM to the air pump (Figure 6.1) [96]. The flow rate of the TuffPro™ air pump is calibrated between 1.70 and 1.89 L/minwith a Challenger Air Calibrator (BGI, Inc., Waltham, MA). The equipment was carried in a satchel custom-made with local *gitenge* fabric for women, and in a small, colorful backpack for children (Figure 6.2). These measurement devices have been approved for and used successfully with adults, school-aged children [97] and children under 5 years of age [98], including in Rwanda during Phase 1 of this project [6].

Field enumerators started the pump and deployed the HAP as the last step during a household visit. The household cook and/or household head was provided with contact information for the field worker and supervisor in the event of a pump malfunction so that it could be repaired or replaced as quickly as possible. Field workers visited the HAP households unannounced on Day 2 to assess compliance and then retrieved the equipment 48 hours after deployment on Day 3. The participants were counseled to either wear the pump constantly or to keep the pump within 1.5 meters throughout the entire 48-hour period.

Physiological measurements in HAP-monitored households. The same households that were selected for HAP also underwent an extensive panel of physiologic measurements to assess blood pressure, carboxyhemoglobin (COHb) concentrations and O_2 saturation (SpO_2). Blood pressure was assessed among main cooks in these households using a blood pressure monitor with cuff (Omron Corp., Kyoto, Japan), and anthropometric measurements were taken and a dietary survey was administered to control for body mass index (BMI) and salt intake. Three blood pressure measurements were taken at each visit, with two minutes at rest between each measurement. COHb, as a proxy for CO exposure,was assessed noninvasively through two methods: pulse oximetry with a RAD-57 instrument (Masimo, Inc., Irvine, CA) and exhaled CO through a MicroCO instrument (CareFusion, Inc., San Diego, CA) [99]. The RAD-57 yields pulse rate, SpO_2 and COHb concentrations in the blood while the MicroCO yields COHb in the blood and exhaled concentrations of carbon monoxide (CO) in parts per million (ppm) [99]. RAD-57 measurements are obtained by clipping the RAD-57 sensor to either the middle or ring finger of the participant's non-dominant hand. Exhaled CO through MicroCO is obtained

by asking the participant to hold her breath for 20 seconds and then blow steadily into the instrument mouthpiece.

Association of household environment and cooking factors and respiratory disease. While all children under 5 years-old who were enrolled in our study were included in all analyses incorporating cough or cold in the last 7 days, children under 2 months-old were excluded from ALRI analyses, since IMCI methods for identifying ALRI cases are not valid in this age group. The univariate relationships for both child with cough or cold in the last 7 days and child with current IMCI-classified ALRI were explored between 10 independent variables describing various household and cooking characteristics: number of persons per sleeping room, stove type used by household (aggregate of all stoves reported by household), number of stoves, main cooking area, number of meals cooked per day, fuel types (aggregate of all fuel types reported by household), use of kerosene lamps, heating of household, roof type (tin vs. clay tile) and age of child. Pairwise differences in 7-day respiratory illness and current ALRI were assessed at each level of each covariate, unadjusted for other covariates, with the lowest prevalence level of each covariate serving as a reference for that variable.

Unadjusted pairwise comparisons were then followed by multivariate logistic regression analysis, using model-adjusted risk (predicted marginal prevalence) to estimate prevalence of respiratory disease and ALRI. Model-adjusted prevalence ratios and 95% confidence intervals were calculated from predicted marginal prevalence. Observations that had a missing value for one or more independent variables included in the model were not included in the analysis. All interaction terms were assessed for inclusion, and

all interaction terms significant at the level of α=0.3 were included in the final model. For pairwise comparisons among levels of variables included in the model, confounding analyses were performed by removing the least significant variable and assessing the effect of the variable removal on the prevalence ratio. Any variables whose removal resulted in a ≥10% change in the effect estimate of any pairwise comparison were retained in the model. Final models for all pairwise comparisons are specified in the supplemental material.

Correlation of concurrent ALRI and diarrhea episodes

In order to determine the likelihood of current ALRI within one week of reported diarrhea, an unadjusted tetrachoric correlation coefficient was calculated alongside estimates modeled by an adjusted bivariate probit model to assess the correlation between the point prevalence of ALRI and 7-day prevalence of diarrhea, assuming a latent normal variable. Identical procedures were used to assess the correlation between 7-day prevalence of child with a cough or cold and 7-day prevalence of diarrhea. ALRI/diarrhea and ARI/diarrhea comorbidities were quantified by the correlation coefficient calculated through both unadjusted and adjusted methods, with -1 indicating a protective relationship, 0 indicating no relationship (or concurrent cases are by chance alone) and 1 indicating perfect concurrence. Given prior evidence that moderate and severe malnutrition are linked to ALRI severity (Chisti, et al. 2009), we assessed the relationship between middle-upper arm circumference (MUAC) and both ALRI and ARI in this population. Direct MUAC measurements were used in unadjusted models to calculate Pearson's correlation coefficients; however, in adjusted bivariate probit models did not converge, so a binomial MUAC outcome was used, where 0=MUAC≤12.5 cm

and 1=MUAC>12.5 cm. These cut-offs have been established by the WHO and UNICEF to indicate the presence of acute malnutrition [101].

Association of household environment and cooking factors with gravimetric $PM_{2.5}$ measurements.

Initial univariate analyses of 24-hour mean PM2.5 concentration obtained from gravimetric mass revealed an exponential distribution for measurements obtained from both primary cooks and children. Data were therefore logarithmically transformed prior to regression analyses and assessed for normality after transformation. The following variables were examined relative to mean 24-hour $PM_{2.5}$ concentration: stove type, number of stoves, main cooking area, number of meals household cooks per day, fuel type, kerosene use, heating of household and roof type. For each level of each covariate, unadjusted means, unadjusted ln(means) and adjusted conditional marginal ln(means) were calculated. Pairwise comparisons of conditional marginal ln(means) were performed between each level of each covariate and the reference level for that covariate by fitting a weighted linear regression model with Taylor series linearization and Zeger robust standard errors to account for between-cluster variance. In order to determine the appropriate reference group for each covariate, the covariate level had to have the lowest gravimetric ln(mean) $PM_{2.5}$ mass and have a sample size that would allow a difference of 15% from the overall mean to be detected.

$PM_{2.5}$ data processing

We calculated 24-hour mean $PM_{2.5}$ concentration using standard EPA guidelines [95]. Filter volume is calculated using the following equation: $V_a = Q_{AVE} * t * 10^{-3}$, where

Q_{AVE}=1.8L/min and t=1440 minutes (total elapsed pump run time). Filter mass is calculated by subtracting the filter pre-weight from the filter post-weight and multiplied by the unit conversion of ug to mg, 10^3, using the following equation: $M_{PM} = (M_f - M_i) * 10^3$, where M_f is the final mass after deployment and M_i is the pre-weight mass. Finally, PM concentration is calculated with the following equation: $[PM] = M_{PM}/V_a$. Based on the distribution of 24-hour mean $PM_{2.5}$ concentration data, appropriate data transformations will be performed to transform $PM_{2.5}$ concentration data for primary cooks and children to a normally distributed variable for regression analyses.

Association $PM_{2.5}$ concentration with COHb, exhaled CO, blood pressure and pulse

Carboxyhemoglobin (COHb) measurements in primary cooks and children were obtained through both a RAD-57 pulse oximeter (Masimo Corporation, Irvine, CA) and through Micro-CO breath CO monitors (CareFusion, San Diego, CA). Heart rate was obtained through pulse oximetry in both primary cooks and children and blood pressure was obtained from primary cooks only. Univariate analyses were performed for all continuous measurements, including COHb, exhaled CO (ppm), pulse, systolic blood pressure (SBP) and diastolic blood pressure (DBP). Appropriate data transformations were performed to approximate a normal distribution prior to analyses. The relationships between PM2.5 and all appropriately transformed COHb, exhaled CO, pulse, SBP and DBP variables were fit to linear regression models with Zeger robust standard errors to account for between-cluster variance.

Association of gravimetric PM2.5 with child health outcomes, cough/cold and ALRI

Unadjusted and adjusted logistic regression analyses will be performed to assess the relationship between health outcomes of interest in children (cough or cold in the last 7 days and current IMCI-identified ALRI). Generalized estimating equations will be applied to interpret the regression coefficients of logarithmically transformed $PM_{2.5}$ concentration. The interquartile range of $PM_{2.5}$ mass was determined between the first and third quartiles, and the proportional increase (q) was determined by dividing the value of the third quartile by the value of the first quartile. Risk ratios were calculated through a weighted robust Poisson model by the following equation: $RR = e^{\beta_{PM2.5}\log(q)}$ which simplifies to $RR = q^{\beta_{PM_{2.5}}}$[102].

Ethical approval. The study protocol, survey instruments and informed consent was reviewed and approved by the Emory University Institutional Review Board (Ref #: 73615), the London School of Hygiene and Tropical Medicine Research Ethics Committee (Ref # 7711), the Rwandan National Ethics Committee (Ref # 1497) and the National Health Research Committee of Rwanda (Ref # NHRC/2014/PROT/0163).

Results

Balance between study arms

A total of 1582 households with 2179 children were enrolled in the overall study between September and December 2014, of which 225 participated in the 48-hour HAP monitoring and more intensive physiological assessment. Intervention and control groups were roughly equivalent on all cooking and stove use exposure indicators and potentially confounding household characteristics and demographic factors, indicating that our randomization procedure resulted in well-balanced study arms (Table 6.1). The majority

of households only had a three-stone fire(s) (74.18, 95%CI: 70.58-77.49), with no other stove types present, followed by households that only had a stationary, built-in *rondereza* stove (95%CI: 10.92-15.40). The remaining 12.81% of households either had no stove (3.39%, 95%CI: 2.37-4.84), a portable clay *imbabura* only (0.72%, 95%CI: 0.32-1.65) or some combination of all three types of stoves. Most households used a designated kitchen area to cook, with 36.09% (95%CI: 31.73-40.69) of households using a kitchen area within the house and 39.76% (95%CI: 35.39-44.29) using a kitchen room detached from the house. The majority of households use wood only (44.70%, 95%CI: 40.52-48.96) or straw, shrubs, grass or an agricultural crop (31.29, 95%CI: 27.98-34.80) or some combination of both fuel types (19.56%, 95%CI: 16.76-22.70). The remaining 4.45% of households either use charcoal only or some combination of wood, straw, shrubs, grass or crop with charcoal (Table 6.1). The mean age of all primary cooks (33.05 years, SD=10.67; intervention: 32.56, SD=10.27) and of all children (control mean: 27.82 months, SD=43.56; intervention: 27.11, SD=15.50) was similar between intervention and control groups.

Prevalence of health conditions

Both ARI and IMCI-identified ALRI prevalence peaked in children who were 6-11 months-old; in this age group, 43.06% (95%CI: 35.79-50.65) were classified as having had a cough or cold in the previous 7 days, while 15.58% (95% CI: 11.24-21.20) were classified as having current ALRI by the IMCI criteria. Overall 7-day ARI prevalence among all children was 30.66% (95%CI: 28.06-33.39) while overall IMCI-identified ALRI prevalence was 6.86% (95%CI: 5.67-8.28). Among children in the oldest age bracket (48-60 months), ALRI prevalence dropped off rapidly (3.26%, 95%CI: 1.22-

8.40). Overall, females were significantly more likely to have both ARI (RR=1.32, 95%CI: 1.14-1.53) and ALRI (RR=1.46, 95%CI: 1.03-2.06), and this association appeared to become more pronounced as age increased (Table 6.2).

Association of household and cooking characteristics with 7-day prevalence of acute respiratory infection

Multivariate analyses revealed that stove and fuel type were both associated with child cough or cold in the last 7 days, after adjusting for relevant confounding factors. Compared to households that had no stove present, households with a single *imbabura* (RR=4.22, 95%CI: 1.11-16.07), a single *rondereza* stove (RR=2.70, 95%CI: 1.29-5.62) and a single three-stone fire (RR=2.91, 95%CI: 1.41-6.01) presented significant excess risk for respiratory infection. Generally, increasing from one stove to two increased the relative risk. Two *rondereza* stoves (RR=3.31, 95%CI: 1.02-10.73), two three-stone fires (RR=3.75, 1.68-8.39), a three-stone fire with *imbabura* (RR=3.58, 95%CI: 1.43-8.96) and a three-stone fire with *rondereza* (RR=2.85, 95%CI: 1.21-6.70) generally had higher risk ratios than any of those stove types alone. Generally, too few households with two *imbabura* stoves or three total stoves were available to infer the relative risk of ARI, indicated by 7-day prevalence of child cough or cold. After adjusting for relevant confounders, no significant excess or reduced risk was observed by fuel type, cooking area, kerosene lamp usage, heating behaviors, roof type or PCA-derived socioeconomic status. Additionally, no association was observed between respiratory infection and number of persons per sleeping room (Table 6.3).

Association of household and cooking characteristics with IMCI-identified acute lower respiratory infection

Due to low overall ALRI prevalence, exposures to various stove types could not be stratified by number of stoves in the household. Unadjusted analyses indicated an increased risk of ALRI among children living in households with two stoves compared to no stoves (RR=3.75, 95%CI: 1.02-13.88) and where cooking areas were in the sitting room (RR=2.45, 95%CI: 1.17-5.16) or outside the house (RR=1.75, 95%CI: 1.02-3.01) compared to children in households where cooking activities occurred in a separate detached kitchen; however, adjusted analyses only indicated a significant association between ALRI and PCA-derived socioeconomic index, with an increased risk associated with the first four quintiles compared to the fifth (wealthiest) quintile (Table 6.4).

Association of household and cooking characteristics with $PM_{2.5}$ mass

Of the 226 households approached for enrollment into the 48-hour $PM_{2.5}$ personal exposure study, HAP equipment was ultimately deployed for 225 primary cooks and 223 children in 56 intervention and 56 control villages, representing 14.2% of all primary cooks and 10.2% of all the children included in this study. A successful deployment was designated for all deployments in which the pump remained running for at least 44 hours of the full 48-hour period (i.e., 22 of 24 total pump running time hours). HAP equipment was deployed and retrieved successfully after 48 hours in 182 (80.9%) of primary cooks but only 123 (55.2%) of children. After removing observations for which less than 44 hours of monitoring were performed, PM2.5 concentration obtained from both the primary cook and child filters was not normally distributed (primary cook: skewness=2.54, kurtosis=7.80, Shapiro-Wilk W= 0.701, $p<0.001$; child: skewness=2.90,

kurtosis=9.83, Shapiro-Wilk W=0.679, p<0.001). After logarithmic transformation of these two variables, the distribution among children did not differ significantly from a normal distribution (Shapiro-Wilk W=0.99, p=0.417, skewness=0.22, kurtosis=0.41) and while statistical testing for concentration data from primary cooks significantly differed from a normal distribution, skewness and kurtosis values indicated that a normal distribution assumption was valid (primary cook: Shapiro-Wilk W=98, p=0.015, skewness=0.02, kurtosis=-0.80). Therefore, only logarithmically transformed $PM_{2.5}$ data were used in these analyses. After exponentiating log-transformed means, mean $PM_{2.5}$ concentration among primary cooks was 95.45 ug/m^3 +/-3.37 and 193.29 ug/m^3 +/- 2.16 for children.

Among primary cooks, personal 24-hour mean $PM_{2.5}$ concentration was significantly higher among those who cooked in the household sitting room vs a separate kitchen (contrasted ln(mean)=1.46+/-0.46, p=0.002) and among those who cooked more than 3 meals per day compared to a single meal (contrasted ln(mean)=0.73+/-0.33, p=0.028). $PM_{2.5}$ was significantly lower among those who lived in households that used kerosene lamps to light the home, compared to those that reported no kerosene use (contrasted ln(mean)=-0.86+/-0.27, p=0.002). No associations were observed between $PM_{2.5}$ exposure among primary cooks and stove type, number of stoves, fuel type or heating of the household (Table 6.5).

Among children, unadjusted analyses revealed that PM2.5 concentration was lowest in households that used a *rondereza* stove; therefore, these households were used as the reference group to determine the association between $PM_{2.5}$ exposure and stove type.

Adjusted analyses revealed that children in households that had no stove (contrasted ln(mean)=0.88+/-0.35, p=0.014) or had an *imbabura* only (contrasted ln(mean)=1.27+/-0.53, p=0.018), a three-stone fire only (contrasted ln(mean)=0.67+/-0.18, p<0.001) or a three-stone fire with an *imbabura* (contrasted ln(mean)=3.36+/-0.74, p<0.001) all had significantly elevated $PM_{2.5}$ exposure compared to children living in households with a *rondereza* only; however, only 4 children in this HAP sub-study lived in households with no stove and only 1 child lived in a household with both a three-stone fire and *imbabura*. Compared to children living in households with a separate kitchen designated as the main cooking area, children living in households where cooking primarily occurred in the household sitting room appeared to have elevated $PM_{2.5}$ exposure (contrasted ln(mean)=0.61+/-0.34, p=0.073), although this relationship was not significant at α=0.05. After examining child $PM_{2.5}$ exposures among children by fuel type used by the household, households using charcoal only appeared to have the lowest exposure. However, since only one child in this sub-study lived in such a household, this was not used as our reference group. Eighty-two children lived in households that used wood only, and this group had the second lowest unadjusted mean $PM_{2.5}$ exposure among children, so this group was used as the reference. Compared to households that used wood only for fuel, households that used a combination of wood and charcoal had significantly reduced $PM_{2.5}$ levels among children (contrasted ln(mean)=-1.65+/-0.46, p<0.001), although only 2 children lived in households using these two fuel sources. Finally, children living in households that heated their homes had elevated $PM_{2.5}$ exposures (contrasted ln(mean)=0.51+/-0.19, p=0.010) and appeared to have elevated exposures in households that used kerosene lamps for lighting (contrasted ln(mean)=0.51+/-0.19, p=0.078), although this association was not significant at α=0.05. Number of stoves and

number of meals cooked per day did not appear to affect child exposure to $PM_{2.5}$ (Table 6.5).

Association of PM2.5 exposure with cough or cold in the previous 7 days and ALRI

Unadjusted analysis does not reveal any direct association between 24-hour $PM_{2.5}$ exposure and either cough or cold in the previous 7 days or current ALRI. After adjusting for child age, gender and socioeconomic status, an interquartile range 2.63-fold increase in log-transformed $PM_{2.5}$ mass (representing an increase from 118.83 ug/m^3 to 311.92 ug/m^3) was not associated with 7-day prevalence of cough or cold, but tripled the risk of ALRI in children (aRR: 2.99. 95%CI: 1.00-9.25) (Table 6.6).

Correlation of concurrent ALRI, 7-day prevalence of ARI, 7-day prevalence of diarrhea and malnutrition

Seven-day prevalence of diarrhea in the last 7 days that fit the WHO case definition was significantly correlated with current IMCI-identified ALRI, both by unadjusted tetrachoric correlation analysis (ρ=0.215 +/-0.009, p<0.001) and bivariate probit analysis adjusted for age, gender, stove type and socioeconomic index (ρ=0.179 +/-0.062, p=0.004). Unadjusted bivariate logistic regression analysis indicated that 27% of the variance in ALRI prevalence can be explained by diarrhea (Table 6.7). Similarly, unadjusted tetrachoric correlation analysis between diarrhea in the previous 7 days and cough or cold in the previous 7 days was significant (ρ=0.189+/-0.006, p<0.001) and for adjusted bivariate probit analysis (ρ=0.166+/-0.044, p<0.001) (Table 6.7). Unadjusted bivariate logistic regression analyses to explore the relationship between 7-day diarrhea and ARI prevalence indicated that 33.07% of the variance in ARI prevalence can be

explained by diarrhea. Strong, significant correlations were also found between middle-upper arm circumference (MUAC) as an indicator for malnutrition and ALRI and ARI. An inverse correlation was found between MUAC measurements and ALRI (ρ= -0.101+/-0.025, p<0.001), but not between MUAC measurements and ARI (ρ= -0.003+/-0.004, p=0.449). When MUAC was dichotomized at the WHO threshold for malnutrition classification (0=(MUAC ≤12.5 cm); 1=(MUAC>12.5 cm), bivariate probit analysis adjusted for age, gender, stove type and socioeconomic index did not yield a significant correlation between ALRI and ARI and the dichotomized indicator for MUAC (Table 6.7). Age-specific crude predicted marginal prevalence of reported cough or cold in the previous 7 days, current ALRI, respondent-defined diarrhea in the previous 7 days and WHO-defined diarrhea in the previous 7 days shows a roughly similar pattern across all diseases, with prevalence peaking prior to one year of age for all conditions. Visually, age-specific and age group-specific prevalence of all four conditions (7-day prevalence of cough/cold, current ALRI, 7-day prevalence of diarrhea fitting the WHO case definition, and 7-day prevalence of diarrhea defined by the respondent) is plotted in Figure 6.3. Prevalence of all four conditions peaks in children who are 6-11 months old and wanes after 2 years of age.

Discussion

Randomization of our intervention and control households resulted in reasonably well-balanced study arms across all exposures of interest and potential confounding factors. Disease outcomes also appeared to be well-balanced, with the exception of ALRI prevalence among 2-5 month-olds, for whom ALRI prevalence was 5.32% (95%CI: 2.21-12.29) in the intervention arm vs 18.86% (95%CI: 11.06-30.28) in the control arm

(p=0.013). Overall prevalence for cough or cold in the previous 7 days and IMCI-identified ALRI had overlapping confidence intervals, indicating that disease prevalence was similar between the two study arms. Both the prevalence of child cough or cold (acute respiratory infection, ARI) in the previous 7 days and current IMCI-identified ALRI peaked in children who were 6-11 months-old during the baseline assessment. Overall, females were significantly more likely to have ARI or IMCI-identified ALRI than males, although this relationship was only defined by age group for ARI due to reduced study power contributed by low overall ALRI prevalence. These results align with previous research that found an excess risk for ALRI among girls [96]. The prevalence ratio defining the excess risk for ARI among females generally increased by age group, with females more than twice as likely to have had ARI in the previous 7 days than males among children 48-60 months-old (RR=2.08, 95%CI: 1.14-3.82) and no significant elevated risk by gender among younger age groups. Among all age groups, females were nearly 50% more likely than males to have ALRI identified by the IMCI method on the day of the survey (RR=1.46, 95%CI: 1.03-2.06). In a separate HAP study in the Gambia, researchers found that girls tended to be carried on their mothers backs at older ages and for longer periods of time than boys, thereby exposing them to more lengthy HAP exposures [103]; if these results apply in Rwanda, this may explain the higher pneumonia-attributable disease burden borne by girls in our population. Together, this indicates that while risk is highest in the youngest age groups, excess risk presented by gender is neutralized when children are younger when mothers will typically have their children nearby or on their backs while cooking, regardless of their gender. The excess risk presented by female children in older age groups is consistent with the fact that they are more likely to remain nearby or indoors while cooking is occurring.

While excess risk for ARI 7-day prevalence is associated with stove type and number, no such association was observed for ALRI. Overall, ALRI prevalence was much lower than ARI prevalence, and the relatively low number of ALRI cases precluded us from stratifying stove type by number of stoves. Relatively low ALRI prevalence also left our study under-powered to detect a 25% difference from overall prevalence (6.86%) and any level of each covariate. Only our PCA-derived socioeconomic index was associated with ALRI, and this is likely due to the natural quintiles that were calculated to create the index; therefore, the population was equally distributed across all levels of this covariate. With between 434-438 children in each quintile, the study was adequately powered to detect a difference in ALRI prevalence between the first four quintiles and the fifth quintile, in which ALRI prevalence among children was 2.59% (95%CI: 1.42-4.68). For children in which ARI, indicated by cough or cold, occurred in the 7 days before the survey, only stove type and stove number were associated with ARI. No association was observed between 7-day prevalence of ALRI and primary cooking location, household crowding (indicated by number of persons per sleeping room), number of meals cooker per day, fuel type, kerosene usage, household heating, roof type or socioeconomic index.

Personal 24-hour mean $PM_{2.5}$ concentration (ug/m^3) was differentially associated with various cooking, household and socioeconomic factors among primary cooks and children. Since only a subset of the overall population was enrolled in the 48-hour personal $PM_{2.5}$ exposure study, and full 44- to 48-hour measurements were only available for about 81% and 55% of primary cooks and children, respectively, data on $PM_{2.5}$ concentrations by stove type could not be stratified by number of stoves. Additionally,

approximately 3.6% of households in this personal exposure sub-study had unknown stove types. Households that did not have an observed or reported stove did not necessarily lack a stove; however, because stove information was not available for these households, these households were categorized together. Lowest $PM_{2.5}$ concentrations were measured in primary cooks where no stove was observed or reported in the house, which was used as the reference group for primary cooks; meanwhile, for children, lowest $PM_{2.5}$ concentrations were measured in households with *rondereza* stoves only. Compared to primary cooks living in households with no reported or observed stove, stove type did not appear to be associated with $PM_{2.5}$ exposure. In children, though, all stove types (with the exception of households with both a three-stone fire and a *rondereza* stove) were associated with significant increased $PM_{2.5}$ concentration. Even children living in households with no observed or reported stove had higher $PM_{2.5}$ exposures than children living in households with *rondereza* stoves, indicating the presence of unreported stoves and/or high community-level exposures. Anecdotally, children can congregate around three-stone fires and even charcoal-burning *imbabura* stoves as they are low to the ground and open on all sides. The fire pit of a *rondereza* stove is higher off the ground, and the *rondereza* stove itself is typically built against a wall, preventing children from gathering around the fire; therefore, $PM_{2.5}$ exposures among children may have been lowest in households using *rondereza* stoves due to the physical characteristics of the stove itself. These exposure data did not correspond directly with disease outcome data, as *rondereza* stoves appeared to present an excess risk of ARI in children. This should not be surprising, as even if an apparent benefit appears to be conferred by one traditional stove type over another, no stove type reduced emissions to a point where a

health benefit would be expected [104]. Generally, the association between $PM_{2.5}$ concentration and number of stoves in the household was statistically indistinguishable.

Mean 24-hour $PM_{2.5}$ concentration was significantly higher in primary cooks who cooked in the household sitting room compared to those who cooked in separate detached kitchens, and this association was nearly significant among children ($p=0.073$). This corresponds with the unadjusted analysis for the association between ALRI and cooking area, in which children had a nearly 2.5-fold risk of ALRI if the main cooking area was designated as the sitting room; however, this association was not significant after controlling for age, gender, socioeconomic status, stove type and household crowding. The sitting room is typically not well-ventilated, and because the sitting room is the main gathering room in the home, there is little reason to for both primary cooks and children to leave the room while cooking; therefore, we can intuitively reason that exposures may be higher in this sub-group. This comports with intervention priorities that have been previously outlined, which emphasize the need to improve ventilation and change cooking behaviors in addition to introducing lower emission stoves [105]. Curiously, primary cooks who cooked outside of the house appeared to have higher exposures that approached significance ($p=0.056$). As ventilation should not be a concern in outdoor spaces, behavioral factors may be responsible for this trend and should be explored more fully. Primary cooks who cooked three or more meals per day had higher PM2.5 exposure than primary cooks who cooked only one meal per day, with the daily longer duration of stove tending being the likely source of excess exposure.

Fuel type did not seem to affect $PM_{2.5}$ concentration, with the exception of children in households using both wood and charcoal for fuel. Compared to households that used wood only for fuel, children in households using wood and charcoal had significantly reduced exposures. Despite the fact that charcoal stoves, such as the *imbabura* stoves used by our study population, are not necessarily clean, previous research has indicated that charcoal can effectively reduce $PM_{2.5}$ emissions prior to transitioning to more advanced improved cookstoves [106]. While households that used charcoal only were enrolled in our overall study, only one primary cook and one child from the HAP exposure sub-study lived in a household that only used charcoal for fuel; therefore, we could not perform any statistical analysis for children in this sub-group.

The use of kerosene lamps in the household appeared to significantly reduce $PM_{2.5}$ concentration in primary cooks while significantly increasing exposure in young children. The reasons behind this contrasting relationship are not immediately clear and indicates a need for more behavioral studies examining how, when and why kerosene lamps are used. Kerosene lamps are generally thought to deleteriously affect air quality enough to confound or even negate the impact of improved biomass burning stoves [87]; however, they are still considered cleaner than biomass burning stoves. The deleterious health consequences of kerosene use are due primarily to NO_2, benzene and toluene emissions (Muller et al., 2003); the reductions in $PM_{2.5}$ observed in primary cooks may be due to the use of kerosene in place of traditional stoves for nighttime heating and lighting. While the relationship between kerosene usage and $PM_{2.5}$ concentration in children approached significance ($p=0.078$), it is not immediately clear why kerosene usage would increase $PM_{2.5}$ exposures. Household heating through use of traditional

stoves did not affect exposure among primary cooks, but did significantly increase exposure among children. This is likely due to more extended periods of biomass burning, although the reason behind the selective effect on children should be explored in more depth. Finally, children living in homes with tin rooves had significantly less $PM_{2.5}$ exposure than children living in households with clay tile rooves. Since clay tile rooves generally tend to have better natural ventilation, it is difficult to determine why tin rooves might be protective. Incomplete or patchy tin rooves and large seams between the exterior walls and roof may explain ventilation characteristics unique to homes with tin rooves, although this should be explored in more depth.

It is worth noting that although different levels of each covariate was associated with a range of $PM_{2.5}$ concentrations, mean personal $PM_{2.5}$ concentrations at each level of each covariate far exceeded the WHO's Air Quality Guidelines. The WHO has established a maximum 24-hour mean of 10 ug/m^3, above which excess mortality can be expected [91]. At 75 ug/m^3 over 24 hours, previous studies and meta-analyses indicate that a 5% increase in short-term mortality can be expected above a 24-hour mean of 25 ug/m^3 [108]. Only 3.3% of children and 20.3% of primary cooks in our study had measured personal 24-hour $PM_{2.5}$ concentrations that met this 25 ug/m^3 threshold; meanwhile, 57.1% of primary cooks and 87.0% of children had personal $PM_{2.5}$ concentrations exceeding this 75 ug/m^3 benchmark. Further demonstrating this link between $PM_{2.5}$ exposure and ALRI, while adjusted analyses measuring the association between $PM_{2.5}$ exposure and ARI and ALRI outcomes did not reveal a significant association between $PM_{2.5}$ and ARI, they did reveal a significant association between $PM_{2.5}$ exposure and ALRI, with an interquartile increase in $PM_{2.5}$ concentrations doubling the odds of ALRI in young children. This

aligns with the general consensus that the largest contributor to child mortality in households with high HAP exposure is ALRI.

Our analyses addressed ALRI, ARI and diarrhea comorbidity in our study children. Unadjusted correlation analysis ($\rho=0.21$, $p<0.001$) and adjusted bivariate probit modeling ($\rho=0.18$, $p=0.004$) indicated a small but positive correlation between diarrhea fitting the WHO case definition in the previous 7 days and current IMCI-identified ALRI, indicating that these two disease outcomes occurred together more than chance alone. A weaker but still significant correlation was also observed for 7-day diarrhea fitting the WHO case definition and 7-day prevalence of ARI (adjusted bivariate probit $\rho=0.17$, $p<0.001$). These findings align with those of previous studies in Nepal, South India [109] and Ghana [110], which have also found increasingly strong associations with increasing disease severity. We also analyzed the association of MUAC as an indicator of child malnutrition with ALRI and ARI. For adjusted bivariate probit analysis, MUAC was dichotomized into two categories: malnourished (MUAC $\leq$12.5 cm) and normal (MUAC>12.5 cm). No significant relationship was observed, although the correlation between MUAC and ALRI approached significance ($\rho=-0.1192$, $p=0.084$), indicating that the likelihood of ALRI may increase as MUAC decreases. A more significant association was observed when applying MUAC as a continuous rather than binary measure, indicating that MUAC may affect both ALRI and ARI, regardless of whether a child is classified as malnourished or not. This does align with a meta-analysis of previous studies linking mortality from pneumonia to malnutrition, which indicated that a significant increase in pneumonia-attributed mortality can be linked to malnutrition [100]. It is worth noting that while one disease (such as diarrhea) may increase the risk

of respiratory infections, as disease comorbidities may be linked to a shared risk factors or associations between risk factors for individual diseases [110]; however, a longitudinal study in Ghana assessing the relationship between cumulative two-week prevalence of diarrhea and severe ALRI found that diarrhea directly contributes to 26% of ALRI cases (Schmidt et al., 2009). A cohort study among our population, in which children are followed longitudinally and assessed for diarrhea and respiratory infections in a prospective study, would allow this relationship to be characterized using analyses that incorporate temporality, allowing amore reliable determination the association between recent diarrhea morbidy and ALRI [109].

Limitations

Pneumonia cases in this study were not radiographically confirmed and relied solely on our field assessments. Despite lengthy training and piloting periods, enumerator assessment of ALRI using IMCI methods is still fairly subjective. Due to logistical and timing constraints, between 12 and 13 enumerators were working in the field during working hours, so consistency between enumerators is difficult to verify. When possible, though, supervisors and managers watched the IMCI assessments, verified diagnoses and identified opportunities for further training across enumerators. Overall, the relationship between ALRI and $PM_{2.5}$ exposure described in this study is consistent with other studies performed in Sub-Saharan Africa and throughout the world [31], so while subjectivity is a concern, we feel reasonably confident in the conclusions of this study.

Conclusion

This baseline study establishes that both study arms are well-balanced on potential measurable confounders and that both ARI and ALRI disease outcomes vary significantly by age and gender. There was strong evidence linking stove type and number to ARI cases and $PM_{2.5}$ concentration to ALRI in young children. The results of this study may be generalizable to other settings, particularly poor rural settings where households predominantly use wood fires for cooking and heating in Sub-Saharan Africa. Our findings suggest that water and sanitation interventions that reduce diarrheal disease may indirectly reduce ALRI prevalence, lending support to programs that couple water and sanitation interventions with interventions targeting household air pollution, like the program our research team is evaluating. Finally, this baseline study suggests that interventions targeting stove type, fuel type and stove number may reduce overall $PM_{2.5}$ exposure and ARI and ALRI prevalence.

Supplemental Material: Interaction and Confounding Assessment
Appendix 6.2: Interaction and confounding assessment

Interaction was assessed by fitting the full model with all potential confounders with all potential interaction terms. The model was subject to stepwise selection, where a variable would have to be significant at α=0.3 in order for it to remain in the model. Models were refit at each successive step, and variables had to be eligible at α=0.35 in order to remain in the model. All potential confounders and exposures remained in the model for the confounding assessment, but potential interaction terms were eliminated. Interaction was also assessed upon removing each variable from the model by assessing the effect of variable removal on effect size estimates.

After assessing potential interaction terms for the association between all model covariates and ARI, no interaction terms appeared to be significant; however, after assessing potential interaction terms for the association between all model covariates and ALRI, the interaction between stove type and cooking location was significant.

After that, weighted logistic regression models with robust standard errors to account for clustering were fit for each exposure of interest. The full models for both disease outcomes were as follows:

Cough or cold in the past 7 days:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_6(\text{roof2}) + \beta_7(\text{stovetyp}) + \beta_8(\text{fueltyp}) + \beta_9(\text{hl1}) + \beta_{10}(\text{hl3}) + \beta_{11}(numeals) + \beta_{12}(\text{su11}) + \varepsilon$$

Where sesi = PCA-derived socioeconomic index, numstove=number of stoves ever used by the household, cd3=gender, cd6=age, pph2=number of persons per sleeping room, roof2=roof type, stovetyp=all stove types used in household, fueltyp=all fuel types used by household, hl1=kerosene used to light the home, hl3=household heats home, numeals=number of meals prepared by household each day and su11= cooking location. Identical entry and retention criteria were used when assessing interaction in models fit to examine the association of these exposures and ALRI identified through applying the IMCI criteria. After performing backwards selection on all model covariates and potential interaction terms, there was significant interaction between stove type and cooking location. The full model defined for all covariates and ALRI was as follows:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_6(\text{roof2}) + \beta_7(\text{stovetyp}) + \beta_8(\text{fueltyp}) + \beta_9(\text{hl1}) + \beta_{10}(\text{hl3}) + \beta_{11}(numeals) + \beta_{12}(\text{su11}) + \omega(stsu11) + \varepsilon$$

Where stsu11= the interaction between stove type and cooking location. Confounding was then assessed for all models. For each exposure of interest, the full model was first fit and predicted marginal risk ratios obtained for each level of each exposure of interest. Covariates were then removed one at a time and the model re-fit. A particular covariate was permanently removed from the model if it did not appear to substantially impact the risk ratio for all levels of a particular exposure; i.e., the ratio of risk ratios must not have differed by more than 10%.

Confounding assessment

1.1. Socioeconomic status and 7-day prevalence of child cough or cold (ARI). After confounding assessment, numeals, fueltyp, roof2, su11, hl1, hl3, pph2 and numstove were removed. The remaining covariates were significant and retained in the model. This left a final model specified as:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_7(\text{stovetyp}) + \varepsilon$$

1.2. Household crowding, indicated by number of persons per sleeping room, and 7-day prevalence of child cough or cold (ARI). The following variables were removed: numeals, su11, sesi, fueltyp, stovetyp, hl3, hl1 and roof2 were removed. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \varepsilon$$

1.3. Stove types used by household and 7-day prevalence of child cough or cold (ARI). The following variables were removed: numeals, su11, sesi, hl3 and hl1. The following showed evidence of confounding: fueltyp, roof2 and pph2, and remaining covariates were significant in the final model. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_6(\text{roof2}) + \beta_7(\text{stovetyp}) + \beta_8(\text{fueltyp}) + \varepsilon$$

1.4. Number of stoves ever used by household and 7-day prevalence of child cough or cold (ARI). The following variables were removed: numeals, sesi, su11, hl1, hl3 and roof2. The following variables showed evidence of confounding: pph2 and fueltyp. Stove type significantly altered the nature of the relationship between stove number and ARI;

therefore, there appeared to be substantial interaction between stove type and stove number. Ultimately, the model outlined in 2.1 was stratified by stove number to further elucidate the effect of stove number on the effect sizes describing the association between stove type and ARI.

1.5. Cooking location and 7-day prevalence of child cough or cold (ARI). The following variables were removed: numeals, sesi, stovetyp, hl3, hl1 and pph2. The following were identified as confounders: fueltyp and roof2, and remaining covariates were significant in the model. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_6(\text{roof2}) + \beta_8(\text{fueltyp}) + \beta_{12}(\text{su11}) + \varepsilon$$

1.6. Number of meals cooked in household daily and 7-day prevalence of child cough or cold (ARI). The following variables were removed: sesi stovetyp fueltyp roof2 hl3 hl1 pph2 su11. This left the fi

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_{11}(numeals) + \varepsilon$$

1.7. Fuel types used by household and 7-day prevalence of child cough or cold (ARI). The following variables were removed: numeals, sesi, su11, hl1, roof2, hl3. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_7(\text{stovetyp}) + \beta_8(\text{fueltyp}) + \varepsilon$$

1.8. Kerosene use by household and 7-day prevalence of child cough or cold (ARI). The following variables were removed: sesi, stovetyp, roof2, numeals, hl3, fueltyp, pph2. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_{12}(\text{su11}) + \varepsilon$$

1.9. Household heating and 7-day prevalence of child cough or cold (ARI). The following variables were removed: sesi, stovetyp, roof2, numeals, hl1, fueltyp, pph2. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_{10}(\text{hl3}) + \varepsilon$$

1.10. Roof type and 7-day prevalence of child cough or cold (ARI). The following variables were removed: numeals, sesi, su11, stovetyp, fueltyp, hl3, pph2, hl1. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_6(\text{roof2}) + \varepsilon$$

2.1. Socioeconomic status and 7-day prevalence of child cough or cold (ARI). After confounding assessment, roof2, stsu11, hl3, numeals and fueltyp were removed. The remaining covariates were significant and retained in the model. This left a final model specified as:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_7(\text{stovetyp}) + \beta_9(\text{hl1}) + \varepsilon$$

2.2. Household crowding, indicated by number of persons per sleeping room, and current ALRI. The following variables were removed: stsu11, stovetyp, roof2, hl3, numeals and hl1 were removed. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_8(\text{fueltyp}) + \beta_{12}(\text{su11}) + \varepsilon$$

2.3. Stove type and current ALRI. The following variables were removed: roof2, hl3 and hl1. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_7(\text{stovetyp}) + \beta_8(\text{fueltyp}) + \beta_{11}(numeals) + \beta_{12}(\text{su11}) + \omega(stsu11) + \varepsilon$$

2.4. Number of stoves ever used by household and current ALRI. The following variables were removed: roof2, hl3 and fueltyp, . This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_7(\text{stovetyp}) + \beta_{10}(\text{hl3}) + \beta_{11}(numeals) + \beta_{12}(\text{su11}) + \omega(stsu11) + \varepsilon$$

2.5. Primary cooking location and current ALRI. The following variables were removed: roof2, hl3, numstove, numeals, fueltyp, hl1. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_7(\text{stovetyp}) + \beta_{12}(\text{su11}) + \omega(stsu11) + \varepsilon$$

2.6. Number of meals cooked in household daily and current ALRI. The following variables were removed: roof2, pph2, stovetyp, hl3, stsu11, cd6, hl1 and hl3. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_8(\text{fueltyp}) + +\beta_{12}(\text{su11}) + \varepsilon$$

2.7. Fuel types used by household and current ALRI. The following variables were removed: roof2, stsu11, hl3 and numstove. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_7(\text{stovetyp}) + \beta_8(\text{fueltyp}) + \beta_9(\text{hl1}) + \beta_{11}(numeals) + \beta_{12}(\text{su11}) + \varepsilon$$

2.8. Kerosene use and current ALRI. The following variables were removed from the model: roof2, pph2, stovetyp, hl3 and numeals. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_8(\text{fueltyp}) + \beta_9(\text{hl1}) + \beta_{12}(\text{su11}) + \omega(stsu11) + \varepsilon$$

2.9. Household heating and current ALRI. The following variables were removed from the model: roof2, stsu11, stovetyp, numeals and hl1. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_5(\text{pph2}) + \beta_8(\text{fueltyp}) + \beta_{10}(\text{hl3}) + \beta_{12}(\text{su11}) + \varepsilon$$

2.10. Roof type and current ALRI. The following variables were removed from the model: pph2, stsu11, stovetyp, hl3, numeals, hl1. This left the final model:

$$logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{sesi}) + \beta_2(\text{numstove}) + \beta_3(\text{cd3}) + \beta_4(\text{cd6}) + \beta_6(\text{roof2}) + \beta_8(\text{fueltyp}) + \beta_{12}(\text{su11}) + \varepsilon$$

Table 6.1. Socioeconomic, demographic and stove characteristics and cooking behaviors of households at baseline (n=1582)

Variable	Level	Intervention Group n (%, 95%CI) (n=789)	Control Group n (%, 95%CI) (n=793)	Total n (%, 95%CI) (n=1582)
Number of stoves used by household	1	622 (79.22, 75.49-82.52)	640 (80.72, 76.79-84.12)	1262 (79.63, 76.73, 82.24)
	2	136 (17.13, 14.09-20.67)	110 (13.97, 11.32-17.13)	246 (16.28, 13.91-18.96)
	3	7 (0.83, 0.36-1.92)	3 (0.36, 0.12-1.10)	10 (0.70, 0.33-1.47)
Stove types used by household	No stove	24 (2.82, 1.66-4.74)	40 (4.95, 3.28-7.42)	64 (3.39, 2.37-4.84)
	Imbabura only	5 (0.60, 0.17-2.13)	9 (1.05, 0.49-2.25)	14 (0.72, 0.32-1.65)
	Rondereza only	105 (12.96, 10.35-16.10)	106 (13.11, 10.47-16.30)	211 (13.00, 10.92-15.40)
	Three-stone fire only	588 (75.16, 70.55-79.26)	564 (71.53, 66.29-76.25)	1152 (74.18, 70.58-77.49)
	Rondereza & *imbabura*	7 (0.81, 0.35-1.83)	7 (0.78, 0.35-1.76)	14 (0.80, 0.42-1.52)
	Three stone fire and *imbabura*	19 (2.21, 1.19-4.06)	23 (2.76, 1.41-5.33)	42 (2.36, 1.47-3.76)
	Three stone fire & *rondereza*	38 (5.09, 3.40-7.56)	43 (5.69, 3.98-8.08)	81 (5.25, 3.88-7.08)
	All three types of stoves	3 (0.36, 0.12-1.10)	1 (0.12, 0.02-0.82)	4 (0.29, 0.10-0.82)
Main cooking area	Inside house-designated kitchen	276 (35.36, 29.96-41.17)	321 (39.83, 34.11-45.84)	597 (36.57,32.28-41.08)
	Inside house-bedroom	110 (13.55, 10.81-16.87)	120 (14.63, 11.69-18.17)	230 (13.84, 11.64-16.39)
	Sitting room	34 (4.41, 3.07-6.29)	18 (2.09, 1.34-3.26)	52 (3.78, 2.76-5.17)
	Entryway to household	4 (0.49, 0.18-1.29)	1 (0.12, 0.02-0.82)	5 (0.39, 0.16-0.96)
	Outside house	42 (5.16, 3.33-7.90)	50 (7.31, 4.80-10.98)	92 (5.74, 4.16-7.85)
	Separate kitchen	322 (40.91, 35.39-46.66)	280 (35.78, 29.87-42.15)	602 (39.53, 35.17-44.06)
Location where primary stove is used	Inside house-designated kitchen	258 (34.71, 10.33-16.58)	304 (39.85, 33.94-46.08)	562 (36.09, 31.73-40.69)
	Inside house-bedroom	101 (13.14, 10.33-16.58)	115 (15.03, 11.97-18.71)	216 (13.65, 11.38-16.29)
	Sitting room	35 (4.87, 3.40-6.94)	19 (2.33, 1.51-3.56)	54 (4.19, 3.06-5.70)
	Entryway to household	4 (0.51, 0.19-1.35)	1 (0.12, 0.02-0.87)	5 (0.41, 0.16-1.01)
	Outside house	41 (5.17, 3.34-7.91)	49 (7.31, 4.75-11.08)	90 (5.74, 4.16-7.87)
	Separate kitchen	309 (41.46, 35.97-47.17)	261 (35.10, 29.04-41.68)	570 (39.76, 35.39-44.29)
Frequency of primary stove use	More than once per day	651 (85.56, 82.28-88.32)	643 (85.57, 82.36-88.28)	1294 (85.56, 83.05-87.76)
	Once per day	101 (12.88, 10.23-16.09)	105 (13.69, 11.00-16.91)	206 (13.10, 10.98-15.55)
	2-6 times per week	8 (0.97, 0.45-2.04)	3 (0.38, 0.12-1.16)	11 (0.81, 0.41-1.59)
	Once per week	2 (0.36, 0.08-1.53)	0	2 (0.26, 0.06-1.12)
	Less than once per week	2 (0.24, 0.06-0.94)	3 (0.36, 0.12-1.11)	5 (0.27, 0.10-0.72)
Number of meals household cooks per day	1	214 (27.54, 23.46-32.02)	231 (28.94, 24.88-33.36)	445 (27.91, 24.70-31.37)
	2	500 (63.05, 58.50-67.37)	507 (64.13, 59.93-68.11)	1007 (63.34, 59.84-66.70)
	3 or more	75 (9.42, 6.83-13.96)	55 (6.94, 4.79-9.95)	130 (8.75, 6.72-12.16)
Fuel type	Charcoal only	7 (0.83, 0.25-2.72)	9 (1.05, 0.49-2.25)	16 (0.89, 0.38-2.08)
	Straw, shrubs, grass or agricultural crop	230 (29.14, 25.03-33.62)	294 (37.11, 32.32-42.18)	524 (31.29, 27.98-34.80)
	Wood only	380 (48.84, 43.49-54.22)	264 (33.47, 28.01-39.41)	644 (44.70, 40.52-48.96)
	Straw, shrubs, grass, agricultural crop and charcoal	3 (0.35, 0.08-1.48)	12 (1.15, 0.69-3.00)	15 (0.64, 0.31-1.32)
	Wood and charcoal	26 (3.01, 1.75-5.13)	23 (2.67, 1.56-4.54)	49 (2.92, 1.90-4.46)
	Straw, shrubs, grass, agricultural crop and wood	143 (17.83, 14.36-21.93)	191 (24.25, 20.43-28.52)	334 (19.56, 16.76-22.70)
Kerosene lamps used in house	Yes	121 (15.24, 12.00-19.15)	126 (16.32, 12.64-20.82)	247 (15.53, 12.91-18.57)
Household heats home	Yes	94 (11.82, 9.01-15.35)	121 (15.01, 11.69-19.08)	215 (12.68, 10.37-15.40)
Number of persons per sleeping room	1-2	130 (27.00, 22.88-31.56)	116 (23.75, 19.70-28.33)	246 (26.13, 22.89-29.66)
	3-4	318 (63.62, 58.81-68.18)	345 (69.19, 64.27-73.71)	663 (65.11, 61.36-68.68)
	5-6	46 (9.01, 6.81-11.84)	32 (6.11, 4.15-8.90)	78 (8.24, 6.50-10.39)
	7-8	2 (0.37, 0.09-1.45)	5 (0.96, 0.41-2.26)	7 (0.53, 0.23-1.19)
Socioeconomic Index	Lowest	165 (21.89, 17.90-26.48)	151 (19.46, 16.41-22.92)	316 (21.24, 18.16-24.67)
	Second	146 (18.41, 15.68-21.48)	171 (21.43, 18.54-24.64)	317 (19.22, 17.05-21.59)
	Middle	146 (18.15, 15.51-21.12)	170 (21.35, 18.56-24.44)	316 (19.01, 16.91-21.31)
	Fourth	158 (19.95, 17.26-22.94)	159 (19.37, 16.48-22.62)	317 (19.79, 17.65-22.12)
	Highest	174 (21.61, 18.24-25.40)	142 (18.38, 15.20-22.06)	316 (20.74, 18.10-23.65)

Table 6.2. Associations between prevalence of reported cough or cold in the previous 7 days and IMCI-identified ALRI with age and gender.

Reported prevalence of cough or cold in the previous 7 days, by intervention arm									
Age and gender group	**Weighted sample size**	**Intervention Arm % (95%CI)**	**Control Arm % (95%CI)**	**Crude total, % (95%CI)**	**Crude prevalence ratio**	**95% CI**	**Model-adjusted predicted marginal %, (95%CI)**	**Model-adjusted predicted marginal prevalence ratio**	**95% CI**
0-5	8614	33.58 (24.53-44.02)	31.79 (23.17-41.87)	33.12 (25.89-41.25)	1.61	1.04-2.48	30.48 (22.62-39.68)	1.52	0.94-2.48
Female	4053	39.42 (25.93-54.74)	30.61 (18.72-45.78)	37.34 (26.48-49.65)	1.27	0.82-1.97	37.34 (26.47-49.67)	1.27	0.82-1.97
Male	4561	28.13 (17.70-41.60)	32.69 (21.68-46.02)	29.37 (20.88-39.59)	REF	REF	29.38 (20.83-39.67)	REF	REF
6-11	10810	40.18 (31.13-49.95)	41.60 (34.41-49.17)	41.20 (34.45-48.31)	2.00	1.39-2.87	43.06 (35.79-50.65)	2.15	1.41-3.28
Female	5504	46.29 (32.39-60.80)	39.50 (29.85-50.04)	44.25 (34.02-54.99)	1.20	0.84-1.73	44.23 (33.99-54.99)	1.20	0.83-1.73
Male	5306	33.71 (22.35-47.33)	43.69 (32.58-55.48)	36.84 (27.95-46.73)	REF	REF	36.86 (27.96-46.76)	REF	REF
12-23	18704	34.00 (27.34-41.37)	31.27 (25.54-37.64)	33.51 (28.34-39.11)	1.63	1.13-2.34	32.28 (26.75-38.35)	1.61	1.08-2.41
Female	10232	36.91 (27.46-47.47)	38.29 (29.58-47.81)	37.25 (29.72-45.46)	1.31	0.95-1.80	37.20 (29.66-45.43)	1.31	0.95-1.80
Male	8472	30.27 (21.91-40.18)	24.14 (17.30-32.62)	28.44 (22.06-35.81)	REF	REF	28.49 (22.09-35.90)	REF	REF
24-35	19726	32.41 (26.47-38.98)	21.29 (15.95-27.81)	29.57 (24.93-34.66)	1.43	1.01-2.03	30.02 (25.29-35.22)	1.50	1.03-2.19
Female	9346	36.93 (27.31-47.70)	20.09 (13.66-28.54)	32.37 (25.04-40.67)	1.21	0.85-1.72	32.40 (25.03-40.75)	1.21	0.85-1.73
Male	10380	28.39 (21.04-37.11)	22.39 (14.67-32.63)	26.81 (20.90-33.69)	REF	REF	26.79 (20.86-33.69)	REF	REF
36-47	22165	27.61 (21.83-34.25)	26.93 (20.95-33.90)	27.59 (22.98-32.74)	1.34	0.91-1.98	29.45 (24.47-34.96)	1.47	0.96-2.25
Female	11448	32.55 (25.37-40.65)	28.57 (20.79-37.86)	31.45 (25.73-37.80)	1.36	1.04-1.78	31.43 (25.70-37.80)	1.36	1.04-1.77
Male	10717	22.46 (15.89-30.76)	25.06 (17.91-33.89)	23.13 (17.76-29.54)	REF	REF	23.15 (17.80-29.53)	REF	REF
48-60	8126	22.72 (14.84-33.16)	15.49 (8.91-25.55)	20.61 (14.50-28.45)	REF	REF	20.00 (13.47-28.66)	REF	REF
Female	4039	30.78 (19.89-44.33)	23.53 (12.75-39.34)	28.90 (20.07-29.70)	2.33	1.25-4.35	28.90 (20.06-39.70)	2.08	1.14-3.82
Male	4087	14.00 (6.63-27.19)	9.11 (3.48-21.80)	12.42 (6.79-21.65)	REF	REF	12.42 (6.79-21.63)	REF	REF
Female	44622	36.59 (31.82-41.63)	30.10 (25.70-34.90)	34.86 (31.15-38.76)	1.32	1.14-1.53	34.87 (31.12-38.83)	1.32	1.14-1.54
Male	43524	26.56 (22.75-30.75)	25.86 (21.81-30.39)	26.40 (23.39-29.65)	REF	REF	26.35 (23.35-29.58)	REF	REF
Total	88146	31.68 (28.35-35.22)	27.95 (24.64-31.52)	30.66 (28.06-33.39)					
IMCI-identified cases of ALRI									
Age and gender group	**Weighted sample size**	**Intervention Arm % (95%CI)**	**Control Arm % (95%CI)**	**Crude total, % (95%CI)**	**Crude prevalence ratio**	**95% CI**	**Model-adjusted predicted marginal, %, (95%CI)**	**Model-adjusted predicted marginal prevalence ratio**	**95% CI**
2-5 months[1]	6781	5.32 (2.21-12.29)	18.86 (11.06-30.28)	8.81 (5.30-14.27)	4.12	1.30-13.00	9.30 (5.23-16.00)	2.85	0.91-9.00
6-11 months	11493	14.55 (9.62-21.40)	15.68 (10.81-22.20)	14.57 (10.73-19.49)	6.81	2.32-20.02	15.58 (11.24-21.20)	4.78	1.72-13.28
12-23 months	20055	10.60 (6.55-16.73)	7.26 (4.48-11.55)	9.84 (6.67-14.29)	4.60	1.52-13.96	11.08 (7.58-15.91)	3.40	1.17-9.86
24-35 months	21315	3.60 (1.87-6.82)	5.96 (3.59-9.73)	4.46 (2.86-6.88)	2.08	0.67-6.51	5.55 (3.68-15.91)	1.70	0.59-4.92
36-47 months	23708	4.29 (2.44-7.42)	3.19 (1.71-5.90)	3.93 (2.47-6.18)	1.83	0.61-5.48	5.04 (3.28-7.65)	1.55	0.52-4.58
48-60 months	8134	2.69 (0.87-8.04)	0.80 (0.11-5.29)	2.14 (0.76-5.85)	REF	REF	3.26 (1.22-8.40)	REF	REF
Female	46293	8.11 (5.90-11.05)	9.01 (6.97-12.06)	8.34 (6.54-10.59)	1.55	1.06-2.26	9.41 (7.53-11.71)	1.46	1.03-2.06
Male	45194	5.19 (3.48-7.66)	5.74 (4.04-8.08)	5.39 (4.00-7.23)	REF	REF	6.47 (4.96-8.40)	REF	REF
Total	91487	6.69 (5.20-8.57)	7.33 (5.86-9.13)	6.86 (5.67-8.28)					

[1]IMCI procedures for identifying ALRI cases (detection of chest indrawing, rapid breathing and stridor) is valid only in children older than 2 months-old. Models assessing the effect of age on cough/cold in previous 7 days were adjusted for stove type, roof type, gender and number of people per sleeping room.
Models assessing the effect of gender on cough/cold in previous 7 days were adjusted for age and stove type.

Table 6.3. Associations between child cough or cold in the previous 7 days, as reported by the survey respondent, and household, demographic and cooking factors.

Primary cook reported child cough or cold in last 7 days						
Variable	**Levels**	**Unweighted cases/ population**	**Wt. crude prevalence % (95%CI)**	**Crude prevalence ratio (95%CI)**	**Wt. model-adjusted predicted marginal % (95%CI)**	**Predicted marginal prevalence ratio (95%CI)**
Number of persons per sleeping room	3-5	190/688	28.85 (24.84-33.22)	0.88 (0.73-1.06)	28.77 (24.67-33.26)	0.89 (0.74-1.07)
	6-8	2/5	38.81 (7.55-83.13)	1.19 (0.34-4.15)	42.69 (9.43-84.19)	1.32 (0.43-4.05)
	1-2	390/1249	32.71 (29.14-36.50)	REF	32.42 (28.94-36.10)	REF
Stove type (1 stove)	*Imbabura* only	7/18	29.48 (11.88-56.45)	2.32 (0.79-6.80)	46.30 (9.19-88.03)	4.22 (1.11-16.07)
	Rondereza only	75/263	30.26 (24.28-36.99)	2.38 (1.23-4.60)	29.57 (23.47-36.50)	2.70 (1.29-5.62)
	Three-stone fire only	433/1438	31.76 (28.46-35.24)	2.50 (1.30-4.80)	31.88 (28.56-35.39)	2.91 (1.41-6.01)
	No stove observed/reported	3/93	12.72 (6.43-23.63)	REF	10.97 (5.13-21.92)	REF
Stove type[1] (2 stoves)	*Rondereza* only	4/16	33.59 (11.02-67.39)	2.64 (0.84-8.27)	35.05 (12.00-68.11)	3.31 (1.02-10.73)
	Three-stone fire only	54/132	40.45 (31.45-50.14)	3.18 (1.58-6.39)	39.70 (29.27-51.16)	3.75 (1.68-8.39)
	Rondereza and *imbabura*	5/19	17.11 (6.06-39.78)	1.35 (0.39-4.58)	20.61 (6.80-48.03)	1.95 (0.53-7.21)
	Three-stone fire and *imbabura*	23/62	34.41 (22.30-48.95)	2.71 (1.27-4.58)	37.87 (22.64-55.95)	3.58 (1.43-8.96)
	Three-stone fire and *rondereza*	28/96	28.84 (20.25-39.29)	2.27 (1.09-5.77)	30.11 (20.19-42.32)	2.85 (1.21-6.70)
	No stove observed/reported	3/93	12.72 (6.43-23.63)	REF	10.58 (4.77-21.83)	REF
Main cooking area[1]	Inside house- kitchen	255/796	33.09 (28.48-38.04)	1.12 (0.91-1.38)	33.13 (28.46-38.16)	1.14 (0.93-1.40)
	Inside house-bedroom	89/305	31.98 (25.79-38.88)	1.09 (0.87-1.35)	32.06 (26.08-38.70)	1.10 (0.89-1.37)
	Sitting room	22/73	30.20 (19.03-44.33)	1.02 (0.66-1.59)	29.60 (19.01-42.98)	1.02 (0.67-1.55)
	Outside house	47/144	30.93 (22.99-40.17)	1.05 (0.77-1.42)	30.15 (22.72-38.78)	1.04 (0.78-1.39)
	Separate kitchen	229/824	29.49 (25.53-33.78)	REF	29.05 (25.15-33.29)	REF
Number of meals household cooks per day	2	409/1373	31.50 (28.35-34.83)	1.03 (0.88-1.20)	31.28 (28.13-34.61)	1.03 (0.87-1.21)
	3	55/177	30.18 (22.16-39.64)	0.98 (0.70-1.38)	30.05 (22.25-39.21)	0.99 (0.71-1.37)
	1	180/604	30.70 (26.55-35.19)	REF	30.46 (26.27-35.00)	REF
Fuel type	Charcoal only	9/23	33.08 (16.16-55.90)	1.29 (0.51-3.30)	48.21 (10.24-88.37)	2.23 (0.63-7.90)
	Straw, shrubs, grass or agricultural crop	210/718	29.23 (24.79-34.10)	1.14 (0.69-1.90)	30.12 (25.41-35.29)	1.39 (0.73-2.68)
	Wood only	243/852	31.79 (28.16-35.66)	1.24 (0.75-2.07)	32.37 (28.64-36.34)	1.50 (0.77-2.91)
	Straw, shrubs, grass, agricultural crop and charcoal	9/23	43.53 (14.53-77.76)	1.70 (0.58-4.98)	37.86 (9.58-77.79)	1.75 (0.54-5.71)
	Straw, shrubs, grass, agricultural crop and wood	150/469	33.11 (27.88-38.80)	1.30 (0.79-2.14)	21.61 (10.95-38.21)	1.46 (0.75-2.82)
	Wood and charcoal	23/69	25.56 (15.28-39.52)	REF	31.48 (26.37-37.08)	REF
Kerosene lamps used in home	Yes	109/347	30.16 (24.83-36.10)	0.96 (0.78-1.19)	28.75 (23.68-34.41)	0.92 (0.74-1.14)
	No	535/1807	31.34 (28.35-34.50)	REF	31.34 (28.37-34.48)	REF
Household heats home	Yes	87/301	28.71 (21.96-36.57)	0.91 (0.70-1.18)	28.05 (21.48-35.72)	0.89 (0.69-1.16)
	No	557/1853	31.53 (28.77-34.43)	REF	31.39 (28.66-34.27)	REF
Roof type	Tin/Metal Roof	249/846	29.81 (25.77-34.19)	0.93 (0.78-1.11)	29.56 (25.53-33.94)	0.93 (0.78-1.10)
	Clay roof	393/1296	32.08 (28.71-35.66)	REF	31.96 (28.68-35.44)	REF
Socioeconomic Index	Lowest	127/430	29.90 (24.19-36.30)	0.98 (0.73-1.30)	29.48 (23.78-35.91)	0.95 (0.70-1.29)
	Second	121/436	29.69 (24.31-35.70)	0.97 (0.73-1.28)	29.61 (24.15-35.73)	0.96 (0.71-1.29)
	Middle	143/434	32.83 (27.60-38.51)	1.07 (0.83-1.39)	32.01 (26.98-37.49)	1.04 (0.79-1.35)
	Fourth	135/427	32.86 (27.23-39.02)	1.07 (0.82-1.40)	33.62 (27.93-39.84)	1.09 (0.83-1.43)
	Highest	118/427	30.66 (24.79-37.23)	REF	30.93 (24.66-37.98)	REF

[1]The following were exposures were not included in this table due to limited sample size: stove types (three stove types of any kind; two *imbabura* stoves) cooking area (entryway to household).

Table 6.4. Associations between current IMCI-identified ALRI (pneumonia) as diagnosed by enumerator and household, demographic and cooking factors.

Current IMCI-identified ALRI case						
Variable	**Levels**	**Unweighted cases/ population**	**Wt. crude prevalence (95%CI)**	**Crude prevalence ratio (95%CI)**	**Wt. model-adjusted predicted marginal prevalence (95%CI)**	**Predicted marginal prevalence ratio (95%CI)**
Number of persons per sleeping room	3-5	48/698	7.28 (5.37-9.80)	1.12 (0.73-1.70)	7.91 (5.95-10.45)	1.01 (0.69-1.47)
	1-2	86/1260	6.52 (4.92-8.60)	REF	7.85 (6.15-9.97)	REF
Stove type used by household[2]	*Rondereza* only	16/287	7.18 (4.14-12.17)	2.71 (0.69-10.64)	7.23 (3.70-13.67)	4.33 (0.39-47.91)
	Three-stone fire only	120/1586	7.02 (5.52-8.88)	2.65 (0.74-9.41)	7.92 (6.12-10.21)	0.63 (0.07-5.84)
	Three stone fire and *imbabura*	5/62	8.37 (4.06-16.44)	3.15 (0.75-13.27)	6.90 (1.25-30.30)	0.60 (0.07-4.94)
	Three stone fire & *rondereza*	10/104	9.63 (5.15-17.30)	3.63 (0.90-14.73)	12.32 (4.66-28.77)	1.07 (0.27-4.23)
	No stove observed/reported	3/93	2.65 (0.75-8.96)	REF	11.54 (1.62-50.83)	REF
Number of stoves in household[2]	**1**	122/1741	6.57 (5.26-8.18)	2.48 (0.70-8.80)	8.05 (5.98-10.75)	3.45 (0.34-34.94)
	2	29/330	9.96 (6.76-14.42)	3.75 (1.02-13.88)	12.82 (7.50-21.07)	5.49 (0.52-58.13)
	0[1]	3/93	2.65 (0.75-8.96)	REF	2.33 (19.91-80.09)	REF
Main cooking area[2]	Inside house- kitchen	61/805	7.19 (5.26-9.77)	1.24 (0.78-1.98)	6.13 (3.58-10.30)	0.59 (0.15-2.25)
	Inside house-bedroom	18/310	6.10 (3.43-10.62)	1.05 (0.54-2.05)	5.94 (3.40-10.19)	0.57 (0.17-1.90)
	Sitting room	9/73	14.19 (7.07-26.47)	2.45 (1.17-5.16)	14.29 (6.81-27.57)	1.37 (0.44-4.24)
	Outside house	14/145	10.15 (6.22-16.14)	1.75 (1.02-3.01)	13.89 (6.25-28.07)	1.33 (0.75-2.38)
	Separate kitchen	52/834	5.79 (4.18-7.96)	REF	10.43 (4.18-23.72)	REF
Number of meals household cooks per day	2	90/1390	6.29 (4.94-7.96)	0.76 (0.52-1.12)	7.28 (5.86-9.01)	0.82 (0.58-1.15)
	3	13/184	7.23 (4.08-12.49)	0.88 (0.45-1.71)	10.11 (5.83-16.96)	1.13 (0.60-2.15)
	1	51/605	8.24 (6.09-11.06)	REF	8.92 (6.77-11.68)	
Fuel type*	Straw, shrubs, grass or agricultural crop	61/721	7.91(5.72-10.86)	1.84 (0.68-5.02)	8.96 (6.29-12.60)	1.13 (0.22-5.77)
	Wood only	38/872	4.67 (3.28-6.60)	1.09 (0.40-2.95)	6.25 (4.20-9.19)	0.78 (0.15-4.13)
	Straw, shrubs, grass, agricultural crop and charcoal	3/23	13.79 (4.91-33.15)	3.21 (0.76-13.49)	15.76 (1.82-65.34)	1.98 (0.47-8.35)
	Straw, shrubs, grass, agricultural crop and wood	49/471	10.84 (7.87-14.74)	2.52 (0.92-6.95)	11.16 (7.81-15.69)	1.40 (0.29-6.71)
	Wood and charcoal	3/69	4.29 (1.62-10.88)	REF	7.96 (1.75-29.57)	REF
Kerosene lamps used in home	Yes	22/350	4.67 (2.86-7.53)	0.64 (0.37-1.09)	5.99 (3.77-9.40)	0.72 (0.44-1.19)
	No	132/1829	7.31 (5.95-8.96)	REF	8.29 (6.89-9.94)	REF
Household heats home	Yes	28/301	9.32 (5.78-14.70)	1.42 (0.86-2.36)	9.84 (6.17-15.34)	1.29 (0.77-2.14)
	No	126/1878	6.56 (5.35-8.02)	REF	7.66 (6.27-9.31)	REF
Roof type	Tin/Metal Roof	61/849	7.06 (5.29-9.38)	1.04 (0.72-1.51)	8.02 (5.93-10.76)	1.01 (0.65-1.56)
	Clay roof	92/1317	6.80 (5.30-8.68)	REF	7.97 (6.16-10.25)	REF
Socioeconomic Index	Lowest	35/434	7.29 (4.72-11.10)	2.82 (1.41-5.63)	8.52 (5.55-12.85)	2.05 (1.09-3.84)
	Second	33/438	7.95 (5.48-11.38)	3.07 (1.51-6.25)	8.24 (5.56-12.04)	1.98 (1.03-3.82)
	Middle	38/437	8.02 (5.40-11.75)	3.10 (1.49-6.45)	8.98 (6.00-13.22)	2.16 (1.11-4.18)
	Fourth	35/434	8.11 (5.43-11.94)	3.14 (1.51-6.51)	10.47 (7.06-15.26)	2.52 (1.31-4.82)
	Highest	13/436	2.59 (1.42-4.68)	REF	4.16 (2.52-6.79)	REF

[1]No stove observed or reported

[2]The following were exposures were not included in this table due to limited sample size: Number of persons per sleeping room (6-8); stove type (*imbabura* only, *rondereza* and *imbabura*, all three stove types); number of stoves per household (three); main cooking area (entryway to household); fuel type (charcoal only)

Table 6.5: PM2.5 concentration (ug/m^3) by exposure level in primary cooks and children, where monitoring occurred for at least 44 of 48 hours.

	Personal PM2.5 concentration (ug/m^3) for primary cooks						Personal PM2.5 concentration (ug/m^3) for children					
Variable	n	Crude –ug/m^3 (+/- SE)	Crude log(mean) (+/- SE)	Adjusted log LS mean (+/- SE)[3]	Contrasted conditional marginal means	T-stat (p-value)	n	Crude mean-ug (+/- SE)	Crude log(mean) (+/- SE)	Adjusted log LS mean (+/- SE)	Contrasted conditional marginal means	T-stat (p-value)
Stove type used by household												
No stove	8	79.84+/-1.40	4.38+/-0.34	4.46+/-0.49	REF	REF	4	311.06+/-1.35	5.74+/-0.30	5.56+/-0.38	0.88+/-0.35	2.49 (0.014)
Rondereza only	22	123.97+/-1.38	4.82+/-0.32	4.87+/-0.25	0.41+/-0.57	0.73 (0.468)	14	123.97+/-1.13	4.82+/-0.12	4.68+/-0.16	REF	REF
Three-stone fire only	130	87.36+/-1.13	4.47+/-0.12	4.41+/-0.12	-0.05+/-0.53	-0.09 (0.931)	88	223.63+/-1.08	5.41+/-0.08	5.35+/-0.07	0.67+/-0.18	3.72 (<0.001)
Three stone fire & *rondereza*	12	135.64+/-1.40	4.91+/-0.34	5.29+/-0.49	0.84+/-0.51	1.65 (0.102)	11	165.67+/-1.17	5.11+/-0.16	5.58+/-0.27	0.02+/-0.59	0.03 (0.979)
Number of stoves												
0 (No stove, stove unobserved or undescribed)	8	79.84+/-1.40	4.38+/-0.34	4.32+/-0.27	REF	REF	4	311.06+/-1.35	5.74+/-0.30	5.18+/-0.15	0.00+/-0.00	-
1	137	93.69+/-1.13	4.54+/-0.12	4.61+/-0.13	0.29+/-0.34	0.85 (0.398)	92	212.72+/-1.08	5.36+/-0.08	5.38+/-0.07	0.20+/-0.19	1.05 (0.297)
2	31	101.49+/-1.21	4.62+/-0.19	4.32+/-0.27	0.00+/-0.00	-	23	174.16+/-1.15	5.16+/-0.14	5.18+/-0.15	REF	REF
Main cooking area												
Inside house-designated kitchen room	60	84.77+/-1.19	4.44+/-0.17	4.54+/-0.32	0.18+/-0.56	0.31(0.753)	40	230.44+/-1.11	5.44+/-0.11	5.07+/-0.21	-0.38+/-0.37	-1.03(0.305)
Inside house-bedroom	27	108.85+/-1.28	4.69+/-0.25	4.71+/-0.29	0.34+/-0.49	0.70(0.485)	16	284.29+/-1.22	5.65+/-0.20	5.33+/-0.31	-0.12+/-0.45	-0.26(0.798)
Sitting room	4	311.06+/-1.34	5.74+/-0.29	5.82+/-0.37	1.46+/-0.46	3.18(0.002)	3	528.48+/-1.68	6.27+/-0.52	6.05+/-0.26	0.61+/-0.34	1.80(0.073)
Outside house	13	125.21+/-1.31	4.83+/-0.27	5.08+/-0.38	0.72+/-0.37	1.92(0.056)	5	188.67+/-1.34	5.24+/-0.29	5.72+/-0.34	0.27+/-0.33	0.82(0.412)
Separate kitchen	71	90.92+/-1.16	4.51+/-0.15	4.37+/-0.28	REF	REF	54	165.67+/-1.12	5.11+/-0.11	5.44+/-0.18	REF	REF
Number of meals household cooks per day												
1	50	84.77+/-1.27	4.44+/-0.24	4.42+/-0.24	REF	REF	32	194.42+/-1.17	5.27+/-0.16	5.26+/-0.14	REF	REF
2	111	94.63+/-1.15	4.55+/-0.14	4.52+/-0.12	0.10+/-0.29	0.35(0.726)	75	221.41+/-1.11	5.40+/-0.10	5.33+/-0.07	0.07+/-0.17	0.41(0.683)
3 or more	16	152.93+/-1.27	5.03+/-0.24	5.15+/-0.24	0.73+/-0.33	2.22(0.028)	12	172.43+/-1.25	5.15+/-0.22	5.56+/-0.24	0.30+/-0.27	1.12(0.27)
Fuel type												
Straw, shrubs, grass or agricultural crop	57	97.51+/-1.22	4.58+/-0.20	4.61+/-0.18	0.08+/-0.21	0.39(0.698)	40	252.14+/-1.15	5.53+/-0.14	5.50+/-0.11	0.20+/-0.15	1.32(0.189)
Wood only	82	96.54+/-1.14	4.57+/-0.13	4.53+/-0.13	REF	REF	57	179.47+/-1.12	5.19+/-0.11	5.30+/-0.09	REF	REF
Wood and charcoal	5	336.97+/-1.62	5.82+/-0.48	5.08+/-0.37	0.56+/-0.39	1.44(0.152)	2	188.67+/-1.17	5.24+/-0.16	3.65+/-0.42	-1.65+/-0.46	-3.61(<0.001)
Straw, shrubs, grass, agricultural crop and wood	31	78.26+/-1.23	4.36+/-0.21	4.48+/-0.20	-0.05+/-0.23	-0.20(0.844)	19	221.41+/-1.16	5.40+/-0.15	5.26+/-0.17	-0.04+/-0.21	-0.18(0.858)
Kerosene lamps used in home												
Yes	18	47.94+/-1.26	3.87+/-0.23	3.79+/-0.24	-0.86+/-0.27	-3.24(0.002)	16	198.34+/-1.23	5.29+/-0.21	5.55+/-0.13	0.25+/-0.14	1.77(0.078)
No	159	103.54+/-1.12	4.64+/-0.11	4.65+/-0.10	REF	REF	103	210.61+/-1.08	5.35+/-0.08	5.30+/-0.06	REF	REF
Household heats home												
Yes	15	112.17+/-1.43	4.72+/-0.36	4.71+/-0.36	0.18+/-0.39	0.46(0.647)	12	395.44+/-1.14	5.98+/-0.13	5.78+/-0.17	0.51+/-0.19	2.62(0.010)
No	162	93.69+/-1.12	4.54+/-0.11	4.53+/-0.10	REF	REF	107	190.57+/-1.07	5.25+/-0.07	5.28+/-0.06	REF	REF
Roof type												
Tin	63	97.51+/-1.16	4.58+/-0.15	4.52+/-0.19	-0.05+/-0.28	-0.17(0.864)	41	165.67+/-1.12	5.11+/-0.11	5.52+/-0.08	-0.48+/-0.17	-2.75(0.007)
Clay tile	112	92.76+/-1.14	4.53+/-0.13	4.57+/-0.15	REF	REF	78	239.85+/-1.09	5.48+/-0.09	5.04+/-0.13	REF	REF

[1]No observations for stove type = "rondereza and imbabura," cooking area="Entryway to household" or cooking area="Entryway to separate kitchen"

[2] The following were exposures were not included in this table due to limited sample size: Stove type (*imbabura* only, three-stone fire and *imbabura*, all three stove types); number of stoves per household (three); main cooking area (entryway to household); fuel type (charcoal only, straw/shrubs/grass/agricultural crop/charcoal)

[3]Analyses accounted for significant interaction between stove type and cooking location and were adjusted for stove type, fuel type, heating and lighting behaviors, roof type, cooking location, number of meals cooked per day, number of stoves in the household.

Table 6.6: Association of log-transformed 24-hour mean personal child $PM_{2.5}$ concentration with upper and lower respiratory tract infection. Adjusted analyses include child age. Relative risks account for interquartile range (q) increase in logarithmically transformed $PM_{2.5}$ concentration of q=2.63, based on an increase in mean $PM_{2.5}$ concentration from 118.83 ug/m^3 to 311.92 ug/m^3.

Disease outcome	Crude			
	β estimate	**p-value**	**RR (q^{β})**	**95% CI**
Cough or cold in previous 7 days	0.098	0.72	1.10	0.66-1.84
IMCI-identified ALRI	0.903	0.09	2.39	0.86-6.63
	Adjusted			
	β estimate	**p-value**	**RR (q^{β})**	**95% CI**
Cough or cold in previous 7 days	0.153	0.60	1.01	0.96-1.07
IMCI-identified ALRI	1.133	0.06	2.99	1.00-9.25

Table 6.7: Correlation between ALRI and diarrhea 7-day prevalence, obtained through crude correlation analysis and adjusted bivariate probit regression models.

	Crude R^2	Crude ρ	Standard error	M-H X^2 (p-value)	Bivariate probit, Adjusted ρ[1]	Standard error	t (p-value)
Diarrhea and ALRI[2]	0.2700	0.2146	0.0088	665.30 (<0.001)	0.1785	0.0617	2.89 (0.004)
ARI and ALRI[2]	0.9993	0.8971	0.0040	14615.16 (<0.001)	0.9091	0.0278	32.66 (<0.001)
Diarrhea and ARI[2]	0.3307	0.1885	0.0063	902.49 (<0.001)	0.1664	0.0436	3.81 (<0.001)
MUAC and ALRI[3]	0.9206	-0.1013 [3]	0.0248	882.55 (p<0.001)	-0.1192	0.0689	-1.73 (0.084)
MUAC and ARI[3]	0.9234	-0.0026	0.0037	0.57 (0.449)	-0.0715	0.0448	-1.60 (0.110)

[1]Adjusted models include the following covariates: gender, age, stove type, fuel type and SES
[2]Crude ρ measured as tetrachoric correlation coefficient (ALRI, ARI and diarrhea are binomial)
[3] Crude ρ measured as Pearson's correlation coefficient (MUAC=continuous, ALRI & ARI =binomial), but bivariate probit models estimated with binomial MUAC indicator (0=MUAC≤12.5, 1=MUAC>12.5). Bivariate probit model did not converge with continuous MUAC indicator.

Figure 6.1: IMCI criteria for diagnosing a child (2 months to 5 years-old) with a cough with severe pneumonia or pneumonia, adapted from the WHO/UNICEF IMCI Handbook (WHO & UNICEF, 2005).

Signs	Classify As	Identify Treatment
• Any general danger sign OR • Chest indrawing OR • Stridor in calm child	SEVERE PNEUMONIA OR VERY SEVERE DISEASE	Child referred to hospital and CHW notified.
• Rapid breathing	PNEUMONIA	CHW notified
• No signs of pneumonia or very severe disease	NO PNEUMONIA: COUGH OR COLD	

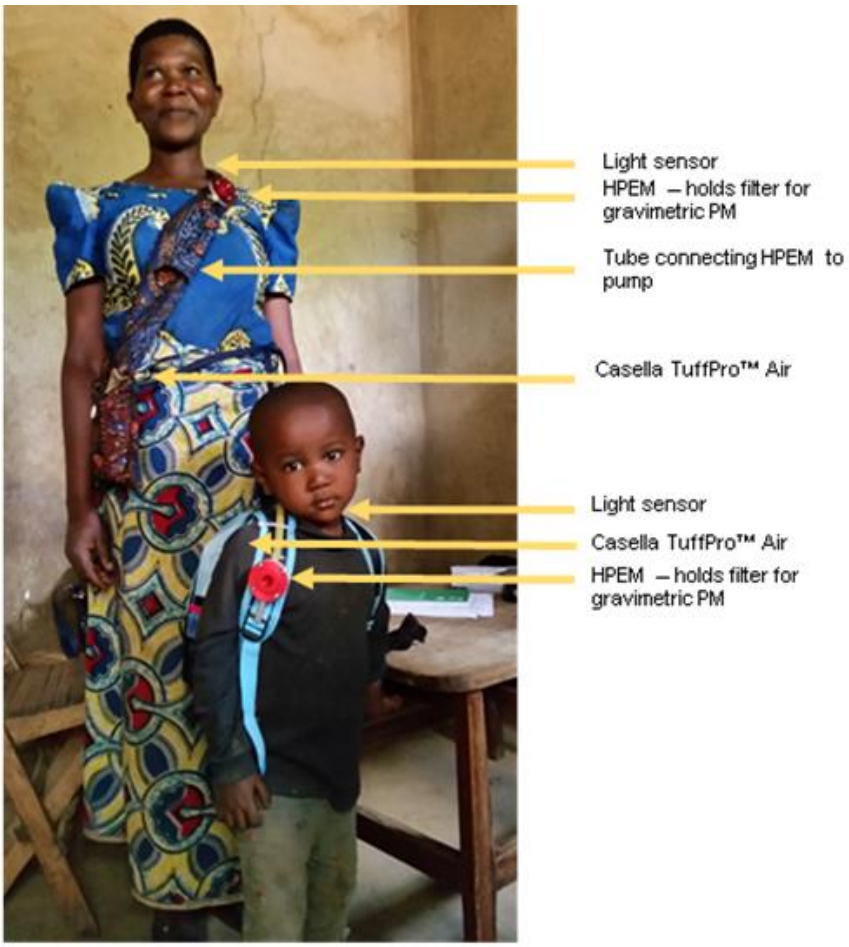
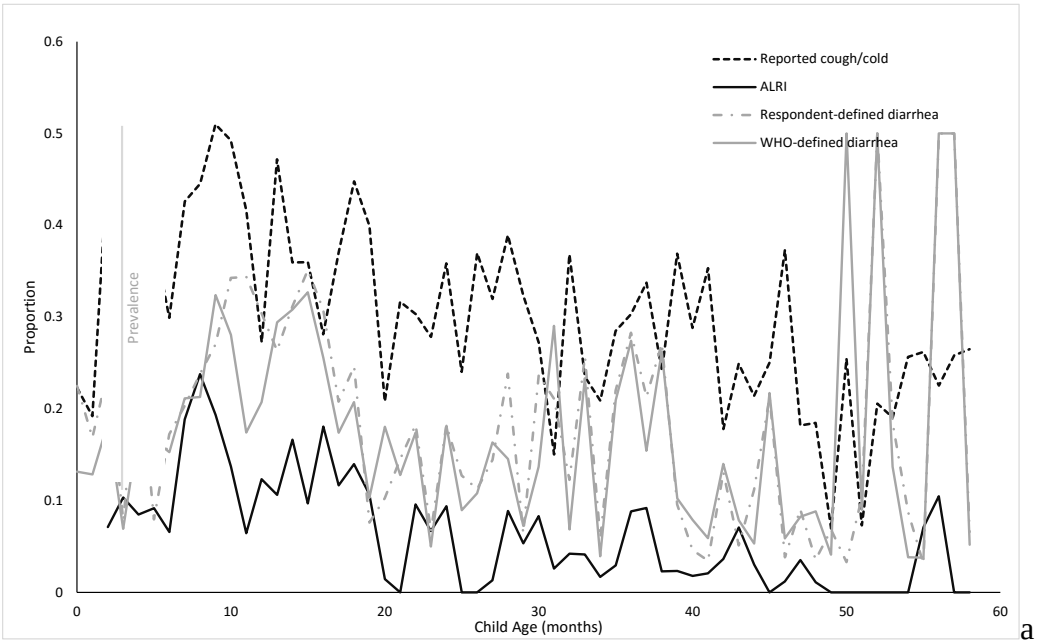
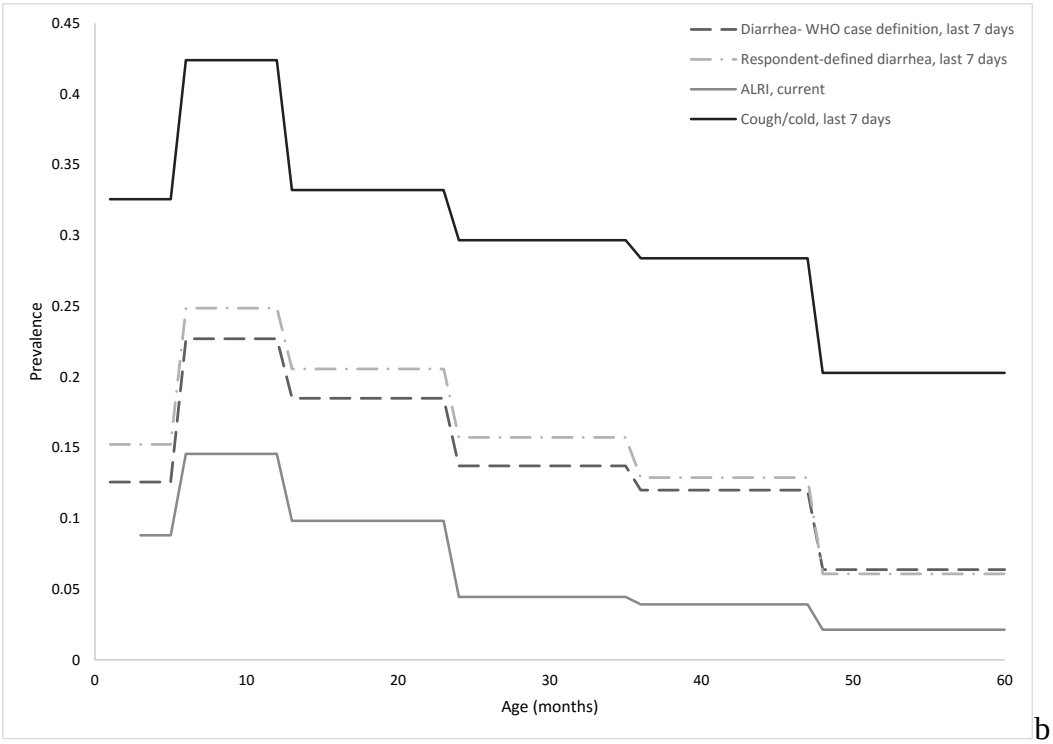


Figure 6.2: HAP equipment set-up for the primary cook and child, showing the configuration of the pump, HPEM with filter, tubes connecting the HPEM to the pump and light sensors. The subjects pictured here are not actual study participants (photo credit: Laura Zambrano).



a



b

Figure 6.3: a) Age-specific crude predicted marginal prevalence of child with reported cough or cold in the previous 7 days, IMCI-identified ALRI, respondent-defined diarrhea in the previous 7 days and diarrhea in the previous 7 days that fit the WHO case definition; b) Age group-specific crude predicted marginal prevalence of child with reported cough or cold in the previous 7 days, IMCI-identified ALRI, respondent-defined diarrhea in the previous 7 days and diarrhea in the previous 7 days that fit the WHO case definition.

Supplemental Material- Principal Components Analysis

Table S6.1: Individual factor loading scores of each variable included in the first principal component from the principal components analysis.

Variable	Coded as:	Factor score
Primary cook education (pced)	0=No formal education or Don't Know or Refused or Missing 1=Nursery 2=Some primary 3=Completed primary 4=Some secondary 5=Completed secondary	0.45798
Household head education (hhed)	0=No formal education or Don't Know or Refused or Missing 1=Nursery 2=Some primary 3=Completed primary 4=Some secondary 5=Completed secondary	0.42990
Electricity (sd4)	0=No 1=Yes	0.68382
Radio (sd5a)	0=No 1=Yes	0.57452
Mobile telephone (sd7a)	0=No 1=Yes	0.72775
Mattress (sd8)	0=No 1=Yes	0.76678
Owns agricultural land (sd13)	0=No 1=Yes	0.16420
Owns home (sd14)	0=No 1=Yes	0.04328
Tends cows (sd15)	0=No 1=Yes	0.18644
Floor type (floor)	0=Mud/earth or Animal Dung or Other 1=Bricks 2=Cement	0.76965
Wall type (wall)	0=Only wood planks or Other 1=Wood planks and mud 2=Mud bricks-not covered 3=Mud bricks- covered with mud 4=Mud bricks-covered with cement 5=Real bricks-not covered 6=Real bricks-covered with cement 7=Wood covered with cement	0.41213
Water source (wu1)	0=Pond/lake/stream/river/rainwater 1=Unprotected dug well/unprotected spring 2=Protected dug well/protected spring 3=Hand pump/borehole/public standpipe 4=Piped water into dwelling/piped water into yard/plot	0.20991
Toilet type (tf2a)	0=Bush/no toilet 1=Pit latrine with slab/pit latrine with no slab/open pit 2=Ventilated pit latrine 3=Composting toilet/flush toilet	0.23667

Supplementary tables: Additional health outcome tables

Table S6.2: Age-specific prevalence of current ARI, severe ALRI cases, respondent-reported child with difficulty breathing, panting or wheezing in last 7 days and respondent-reported child with constant cough in the last 7 days, by intervention arm.

Age Group	EZ + LS Arm N, % (95%CI)	Control Arm N, % (95%CI)	Crude total N, % (95%CI)
Point prevalence of ARI			
0-5 months	30 (27.23, 18.84-37.61)	28 (27.11, 19.05-37.02)	58 (27.20, 20.44-35.20)
6-11 months	40 (32.02, 23.32-42.18)	48 (30.56, 23.89-38.16)	88 (31.58, 25.05-38.91)
12-23 months	52 (24.40, 18.06-32.10)	46 (18.87, 14.11-24.76)	98 (22.92, 18.00-28.71)
24-35 months	42 (16.90, 12.28-22.81)	33 (12.73, 8.90-17.87)	75 (15.82, 12.17-20.32)
36-47 months	53 (19.34, 14.98-24.61)	47 (19.01, 13.63-25.88)	100 (19.26, 15.66-23.45)
48-60 months	19 (19.42, 12.15-29.57)	9 (8.78, 4.40-16.76)	28 (16.36, 10.88-23.87)
Total	236 (22.12, 19.04-25.54)	211 (19.02, 16.46-21.89)	447 (21.29, 18.91-23.88)
IMCI-identified cases of severe ALRI			
2-5 months[1]	4 (4.19, 1.57-10.71)	4 (5.28, 1.99-13.24)	8 (4.47, 2.12-9.16)
6-11 months	3 (2.37, 0.72-7.50))	5 (2.62, 1.09-6.16)	8 (2.44, 1.04-5.61)
12-23 months	2 (0.83, 0.21-3.22)	1 (0.39, 0.05-2.72)	3 (0.71, 0.21-2.36)
24-35 months	2 (0.92, 0.21-3.97)	4 (1.37, 0.51-3.62)	6 (1.03, 0.37-2.88)
36-47 months	0	0	0
48-60 months	0	0	0
Total	11 (1.00, 0.53-1.86)	14 (1.13, 0.69-1.86)	25 (1.03, 0.65-1.64)
Respondent-reported child with difficulty breathing, panting or wheezing			
0-5 months	12 (10.74, 6.10-18.21)	10 (10.32, 5.29-19.16)	22 (10.63, 6.75-16.35)
6-11 months	11 (9.15, 4.88-16.51)	17 (9.74, 5.90-15.68)	28 (9.33, 5.92-14.40)
12-23 months	24 (10.51, 7.02-15.44)	14 (6.49, 3.77-10.96)	38 (9.43, 6.68-13.15)
24-35 months	15 (6.21, 3.69-10.27)	14 (5.71, 3.46-9.29)	29 (6.08, 4.03-9.09)
36-47 months	12 (3.87, 2.13-6.92)	19 (6.57, 4.09-10.39)	31 (4.57, 3.03-6.85)
48-60 months	5 (4.73, 1.73-12.26)	4 (3.36, 1.29-8.45)	9 (4.34, 1.94-9.41)
Total	79 (7.19, 5.52-9.32)	78 (6.85, 5.16-9.06)	157 (7.10, 5.76-8.72)
Respondent-reported child with constant cough in last 7 days			
0-5 months	14 (12.81, 6.94-22.44)	9 (8.87, 4.78-15.85)	23 (11.81, 7.11-18.97)
6-11 months	22 (17.42, 11.07-26.32)	16 (10.16, 6.37-15.82)	38 (15.20, 10.50-21.50)
12-23 months	18 (9.85, 5.70-16.50)	11 (4.68, 2.20-9.68)	29 (8.46, 5.25-13.37)
24-35 months	23 (9.43, 6.23-14.02)	11 (4.42, 2.37-8.11)	34 (8.14, 5.64-11.61)
36-47 months	24 (8.61, 5.31-13.64)	18 (6.47, 3.80-10.81)	42 (8.05, 5.42-11.80)
48-60 months	6 (6.03, 2.69-12.96)	2 (1.92, 0.47-7.50)	8 (4.85, 2.34-9.76)
Total	107 (10.29, 8.07-13.02)	67 (5.96, 4.43-7.98)	174 (9.13, 7.43-11.16)

Correlation between PM2.5 concentration and CO exposure and adult blood pressure
Unadjusted analyses examining the relationship between $PM_{2.5}$ and COHb determined through pulse oximetry among primary cooks revealed a weak but significant negative correlation between $PM_{2.5}$ concentration and %COHb (R^2=0.036, ρ=-0.118, p=0.05). After adjusting for heating and lighting behaviors and number of stoves in the household, though, this association was not significant. No other significant correlations were present among primary cooks between personal PM2.5 exposure and exhaled CO (ppm), pulse, %COHb (as determined through exhaled breath), pulse or blood pressure. Similarly, among children, no significant correlation between PM2.5 exposure and pulse oximetry-based %COHb, exhaled breath %COHb, exhaled CO or pulse was detected.

Table S6.3: Correlation between 48-hour personal PM2.5 exposure and COHb, exhaled CO and pulse among primary cooks and children and blood pressure in adults.

	Unadjusted				Adjusted			
Variable	n	R^2	Coefficient (95%CI)	p-value	n	R^2	Coefficient (95%CI)	p-value
Primary cooks								
ln(COHb), RAD-57	139	0.036	-0.118 (-0.236 - -0.0002)	0.05	139	0.065	-0.097 (-0.212-0.018)	0.10
Pulse, bpm	171	0.014	1.253 (-0.673-3.178)	0.20	171	0.066	1.287 (-0.646-3.220)	0.19
ln(exhaled CO, ppm) Micro-CO	164	0.006	0.047 (-0.058-0.152)	0.38	164	0.051	0.044 (-0.060-0.148)	0.40
ln(COHb), Micro-CO	163	0.006	0.051 (-0.065-0.167)	0.38	161	0.053	0.051 (-0.067-0.169)	0.40
ln(median systolic blood pressure, mmHg)	177	0.002	0.004 (-0.014-0.022)	0.63	175	0.134	0.009 (-0.011-0.029)	0.37
ln(median diastolic blood pressure, mmHg)	177	0.000	0.001 (-0.016-0.017)	0.93	175	0.029	0.002 (-0.015-0.018)	0.86
Children								
ln(COHb), RAD-57	84	0.000	0.022 (-0.277-0.321)	0.88	84	0.112	-0.081 (-0.369-0.207)	0.58
Pulse, bpm	106	0.000	-0.338 (-4.212-3.536)	0.86	106	0.079	-0.862 (-4.757-3.032)	0.66
ln(exhaled CO, ppm) Micro-CO	74	0.002	-0.053 (-0.395-0.290_	0.76	74	0.110	0.012 (-0.387-0.411)	0.95
ln(COHb), Micro-CO	74	0.002	-0.061 (-0.488-0.365)	0.78	74	0.126	0.082 (-0.346-0.540)	0.72

Models included in in Table S6.3:

Primary cook systolic blood pressure:

$$\ln(SBP) = \beta_0 + \beta_1(\text{logcmass}) + \beta_2(\text{hl1}) + \beta_3(hl3) + \beta_4(numstove) + \varepsilon$$

Primary cook diastolic blood pressure:

$$\ln(DBP) = \beta_0 + \beta_1(\text{logcmass}) + \beta_2(\text{hl1}) + \beta_3(hl3) + \beta_4(numstove) + \varepsilon$$

Primary cook carboxyhemoglobin, measured through pulse oximetry:

$$\ln(U9CO) = \beta_0 + \beta_1(\text{logcmass}) + \beta_2(\text{hl1}) + \beta_3(hl3) + \beta_4(numstove) + \varepsilon$$

Primary cook pulse:

$$U8Pulse = \beta_0 + \beta_1(\text{logcmass}) + \beta_2(stovetyp) + \beta_3(su11) + \beta_4(numeals) + \beta_5(numstoves) + \omega(stsu11) + \varepsilon$$

Primary cook exhaled CO (ppm):

$$\ln(wPPM) = \beta_0 + \beta_1(\text{logcmass}) + \beta_2(stovetyp) + \beta_3(fueltyp) + \beta_4(su11) + \beta_5(numeals) + \omega(stsu11) + \varepsilon$$

Primary cook carboxyhemoglobin, measured through Micro-CO:

$$\ln(wCOHb) = \beta_0 + \beta_1(\text{logcmass}) + \beta_2(stovetyp) + \beta_3(fueltyp) + \beta_4(roof2) + \beta_5(su11) + \beta_6(numeals) + \omega(stsu11) + \varepsilon$$

Child carboxyhemoglobin, measured through pulse oximetry:

$$\ln(V9CO) = \beta_0 + \beta_1(\text{logemass}) + \beta_2(stovetyp) + \beta_3(hl3) + \beta_4(su11) + \beta_5(numeals) + \beta_6(numstoves) + \omega(stsu11) + \varepsilon$$

Child Pulse

$$V8Pulse = \beta_0 + \beta_1(\text{logemass}) + \beta_2(stovetyp) + \beta_3(su11) + \beta_4(fueltyp) + \beta_5(hl1) + \omega(stsu11) + \varepsilon$$

Child exhaled CO (ppm)

$$\ln(xPPM) = \beta_0 + \beta_1(\text{logemass}) + \beta_2(stovetyp) + \beta_3(su11) + \beta_4(roof2) + \beta_5(hl1) + \beta_6(numstove) + \omega(stsu11) + \varepsilon$$

Child carboxyhemoglobin, measured through Micro-CO

$$\ln(xCOHb) = \beta_0 + \beta_1(\text{logemass}) + \beta_2(stovetyp) + \beta_3(fueltyp) + \beta_4(hl1) + \beta_5(hl3) + \beta_6(numstove) + \beta_7(su11) + \omega(stsu11) + \varepsilon$$

Chapter 7 - Assessing seroconversion against enteropathogens relative to reported diarrhea and the receipt of a point-of-use water filter in Western Province, Rwanda

Abstract

Diarrhea is a leading contributor to childhood morbidity and mortality in Sub-Saharan Africa. Given the infeasibility of blinding most water, sanitation and hygiene (WASH) interventions, diarrheal disease outcome measures in WASH intervention trials are fraught with potential bias and misclassification. We used the platform of a cluster-randomized controlled trial of a household-based drinking water filter in Western Province, Rwanda to examine the application of enteric seroconversion as an alternative and more objective outcome measure of current and recent infection. All children ≥ 6 and ≤12 months-old among 1582 study households were eligible for enrollment. All enrolled children had their blood drawn through capillary blood draw at baseline and 6 to 9 months after intervention distribution. Multiplex serologic assays for *Giardia*, *Cryptosporidium*, *Entamoeba histolytica*, *Salmonella*, norovirus, *Campylobacter*, enterotoxigenic *E. coli* and *V. cholerae* were performed. The water filter was associated with a significant decrease in *Cryptosporidium* seroconversion. Serologic responses against both *Giardia* and *Cryptosporidium* were positively associated with reported diarrhea in the previous seven days. Children seroconverted against *Cryptosporidium* at relatively early ages (<6 months-old) while *Giardia* seroconversion typically occurred after 12 months. Serological responses for other antigens increased steadily after 6 months of age, plateauing after 12 months. Enteric seroconversion, particularly against protozoa and norovirus, appears to be a suitable objective outcome measure for WASH trials among

children in this age group. Serological reactions against both *Giardia* and *Cryptosporidium* antigens appear to be an appropriate indicator of recent diarrheal disease, and seroconversion against *Cryptosporidium* may be useful as an objective indicator of the effectiveness of WASH interventions.

Introduction

Diarrheal disease is among the top contributors to global child morbidity and mortality (Naghavi et al., 2015; Vos et al., 2015), principally due to inadequate access to clean water and sanitation. Young children ages 0-24 months are particularly vulnerable, given that diarrhea is associated with malnutrition and growth faltering [112], and that poor nutritional status places young children at a high risk of death (O'Neill et al., 2012). Household water treatment (HWT) interventions are more preventive against diarrhea than interventions at the water source, and among HWT interventions, water filtration appears to be the most consistent and effective [114].

Reported diarrheal disease is often used as the disease outcome in water, sanitation and hygiene (WASH) intervention trials, in which caregivers and survey respondents are asked by field enumerators about previous or current diarrhea among children in the household. Such methods are subject to both recall and courtesy biases, are highly subjective and are unable to distinguish specific causal pathogens of disease [115]. Objective measures of exposures or disease are particularly important in trials that cannot be blinded, which is a common methodological issue in environmental health interventions (Clasen & Boisson, 2015). In addition, characterizing age-specific cumulative exposures to enteropathogens can enhance epidemiological surveillance and

inform etiology-specific interventions and regionally-specific treatment methodologies [116].

Stool assays offer opportunity for more objective assessment, but logistical constraints related to the storage, transport and extensive lab work involved can limit its utility in resource-limited settings [26]. Additionally, sensitivity of stool assays to capture etiologic agents can be compromised by pathogens, such as protozoa, that are not continuously shed [117], and prevalence of current cases can vary widely by season. Serological assays that quantify serological responses against various enteropathogens can provide a useful measure of cumulative exposures to enteropathogens in children and supplements objective information to caregiver-reported diarrhea [19]. Microsphere-based multiplex immunoassay methods allow for simultaneous measurement of antibodies against multiple antigens. Previously, this technology has been applied to neglected tropical disease (NTD) surveillance, particularly in light of recent global elimination targets that have been set for various NTDs [20]. In recent years, single enzyme-linked immunosorbent assays (ELISAs) and multiplex immunoassays have been shown to effectively target antigens of enteropathogens (Crump et al., 2007; Moss et al., 2011, 2014a; Priest et al., 2006; Priest et al., 2001, 2010; Steinberg et al., 2004), but the technique's incorporation into intervention trials has not been adequately explored. Trials that have incorporated this approach have not included the full range of enteric antigens that are now available.

Antibody responses to many enteric antigens are not constitutively expressed, allowing researchers to infer some degree of temporality with regard to the timing of infection.

For protozoan infections, cyst-positive stools appear to co-occur with serological responses to *Entamoeba* and *Giardia*, with higher seroprevalence in children for whom stool samples were collected within one week of serum collection versus two weeks or more [118]. This indicates that serological responses can be an indicator of recent protozoan infections. Antibody responses in cryptosporidiosis patients are consistently directed against 27- and 17-kDa *Cryptosporidium* antigens (Cp27 and Cp17, respectively) and are known to have a 12-week half-life (Priest et al., 2001). This study will examine the association between the intervention, a LifeStraw Family™ 2.1 point-of-use water filter, and seroconversion against various viral, bacterial and protozoan enteropathogens among young children as objective measures of exposure to diarrhea-causing agents.

Methods

Intervention. In an effort to reduce the high prevalence of both diarrheal disease and acute lower respiratory infection (ALRI), DelAgua Health, Inc. (DelAgua) distributed EcoZoom Dura™ improved cookstoves and point-of-use Vestergaard-Frandsen Lifestraw Family™ 2.1 advanced water filtration units to 100,000 *Ubudehe* 1 and 2 households, representing the poorest tertile of the population in Western Province. The LifeStraw filter consists of 20 micron filter that water is passed through prior to being filtered again through a 0.2 micron hollow-fiber ultrafiltration membrane into a 5.5L safe storage container. This system can filter up to 18,000L of water, supplying a family of five with clean drinking water for 3 to 5 years. This system exceeds the World Health Organization's "highly protective" standard for household water treatment technologies (Clasen et al., 2009).

The two-pronged cluster-randomized controlled trial design. We assessed the effectiveness of the intervention by conducting a randomized, controlled field trial. Pilot studies in Rwanda on both the LifeStraw™ water filter and EcoZoom™ cookstove have been previously described [5,6], and conservative interpretations of these results were used to design the overall CRT in conjunction with government-reported disease prevalence data [12]. The CRT's design and implementation have been described in detail elsewhere (Nagel et al., 2016 (submitted)). Western Province is a rugged, mountainous and mostly rural region and lies at elevations between 900M in its southern plain up to 3000M in the volcanic foothills of the north. The paired intervention was randomized at a 3:1 ratio by administrative sector, which represented health clinic catchment areas. Ultimately, DelAgua intended to reach all *Ubudehe* 1 and 2 households in 72 sectors with the paired intervention, while 24 sectors are serving as controls until completion of all survey activities in March 2016. Out of a total of ~3600 villages in the province, 174 villages were randomly selected using population proportional selection, divided evenly between intervention and control sectors. We anticipated enrolling a maximum of 10 households with at least one child under-4 years-old per village. We ultimately enrolled 1582 total households with 2179 children. Data collection occurred throughout Western Province, Rwanda with a baseline assessment that was performed from late August through early December 2014. Three follow-up rounds were conducted in the same households from February 2015 through March 2016.

Nested enteric serological study. For this seroconversion sub-study, our enumerators were instructed to enroll all 6 to 12 month-old children they encountered in the RCT

study households with the goal of reaching one 6 to 12 month-old child per village. These children had one blood sample drawn at baseline and one approximately 6 to 9 months later, during the 2nd round of RCT follow-up. Given the exploratory nature of this study in this region, we conservatively estimated that 40% of children would seroconvert from negative to positive for *C. parvum* and norovirus antibody between baseline and follow-up in the absence of any intervention, based on verbal report and a previous trial in Guatemala [19]. If 87 children from each of our two study arms were enrolled in the study, we would have 80% power at α=0.05 to detect a difference in seroconversion of 40% vs. 22.5% in our intervention and control groups.

Before the blood draw, the skin around the area of the blood draw was sanitized with an alcohol pad and allowed to dry. Between 3 to 6 small hanging drops (10uL each for a total of 30-60 uL) from either the child's heel using a 1.00mm x 2.50mm infant Quikheel Microtainer™ lancet (Becton-Dickinson, Franklin Lakes, NJ) or from the child's ring or middle finger using a 1.8mm 21-gauge Microtainer™ finger lancet (Becton-Dickinson, Franklin Lakes NJ), depending on the child's size [122]. During the baseline assessment, all children in this sub-study had their blood drawn either by heelstick or by fingerstick, while all children during follow-up had their blood drawn through fingerstick. The hanging drops of blood were collected on TropBio™ filter discs (Cellabs Pty Ltd, Brookvale, NSW Australia) and kept in individual plastic resealable containers during the fieldwork. Immediately upon return to the field office each day, the discs were placed on a table and allowed to dry overnight. The following morning, they were individually packaged in plastic resealable bags with dessicant, and were sent to the Rwanda National

Reference Laboratory in Kigali, Rwanda for long-term storage at -20°C with 7 days of collection.

Nutritional status was assessed in all children through middle-upper arm circumference (MUAC). MUAC has the same ability to predict short-term mortality as weight-for-height *z* scores [18]. Children 6-59 months-old are classified as being severely malnourished if MUAC ≤11.5 cm, as malnourished if MUAC >11.5 but ≤12.5, and as normal if MUAC >12.5 based on WHO standards [123]. Diarrhea was assessed by asking the respondents to report diarrhea in all household children under-5 in the previous week, both by their own definition and then by the WHO case definition, defined as ≥3 loose stools over a 24 hour period.

Laboratory methods

All laboratory analyses were performed at the Centers for Disease Control and Prevention (CDC) Infectious Disease Laboratories in Atlanta, GA. Total IgG/IgG4 responses against relevant enteropathogens were quantified using a multiplex SeroMAP™ (Luminex Corp, Austin, TX) microsphere-based immunoassay on the Luminex xMAP platform (Luminex Corp, Austin, TX). Antibodies to the following antigens were screened through this assay: *Toxoplasma gondii* surface antigen 2 gene (SAG2); *Giardia* spp. variant-specific protein-3 (VSP3); *Giardia* spp. variant-specific protein-5 (VSP5); *Salmonella* group B lipopolysaccharide, extracted by 3-[(3-Cholamidopropyl) dimethylammonio] propanesulfonic acid (LPS-B CHAPS); *Salmonella* group D, extracted by CHAPS (LPS-D CHAPS); Envelope proteins for three norovirus strains (Norwalk, Sydney and St. Cloud); *Campylobacter* p39 antigen; *Campylobacter* p18 antigen; Enterotoxigenic *E.*

coli (ETEC) heat-labile toxin β subunit (EtxB); *Cryptosporidium parvum* 60S acidic ribosomal protein P2 (CpP2); *Cryptosporidium parvum* 17-kDa protein (Cp17); *Cryptosporidium parvum* 23-kDa protein (Cp23); Cholera toxin β subunit (CtxB); and *Entamoeba histolytica* Gal/GalNAc lectin heavy chain subunit (LecA).

Bead coupling. Procedures describing the coupling process of antigens to microspheres have been described in detail elsewhere [28,124]. The carboxyl groups on each bead were esterified and then reacted with the primary amine groups of each antigen to bind the antigens to the microspheres through a covalent amide bond. 120 ug of each antigen was coupled with 12.5 X 10^6 beads. In addition to the antigens of interest, a control coupling with glutathione-S-transferase (GST) was performed to be used as a background control. Beads were quantified by a hemocytometer and stored at 4°C in phosphate-buffered saline (PBS) with 1.0% bovine serum albumin (BSA), 0.05% Tween 20 and 0.02% sodium azide (NaN_3).

Serum preparation. An *E. coli* extract solution was first prepared and GST production induced to bind excess anti-GST antibodies in serum, which may lead to high background signals. 500 mL of recombinant pGEX4T2/HB101 cells were cultured and GST production was induced by adding 0.5mL of 0.1M isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD reading of ~0.7 absorbance units. The solution was incubated for 1-2 hours at 37°C at 225 RPM. Cells were pelleted, harvested and resuspended with 10mL cold PBS. The *E. coli* extract was added to an extraction buffer with final concentrations of10mM ethylenediaminetetraacetic acid (EDTA), 1mM phenylmethanesulfonyl fluoride (PMSF) protease inhibitor, 1mM *N*-ethylmaleimide

(NEM) and 0.10% NaN_3. The *E. coli* cells were pelleted again and ultimately diluted to a stock concentration of 0.3ug/mL in PBS-based dilution buffer, consisting of 0.50% poly(vinyl alcohol) (PVA),0.80% poly(vinyl pyrrolidone) (PVP), 0.50% casein, 0.30% Tween-20 and 0.02% NaN_3. PVA and PVP were added to reduce background while not affecting specificity [19]. The elution process loosely followed a protocol described elsewhere for antibody elution from DBS [125]. Elution buffer was made with 0.05% Tween-20 and 0.05% NaN_3 in PBS. DBSS were removed from -20°C and brought to room temperature. DBS were considered "complete" if roughly 90% or more of the spot was filled, indicating that approximately 10uL of serum was on the spot. 200uL of elution buffer was aliquoted into microfuge tubes and each spot was eluted in the elution buffer for a minimum of 18 hours at 4°C. The following day, 50uL of eluted serum was aliquoted into 450uL of elution buffer, for a final serum dilution of 1:400.

Multiplex bead assay. Prior to performing the assay, stocks of assay buffer were prepared consisting of 0.50% bovine serum albumin (BSA), 0.02% NaN3 and 0.05% of Tween-20 in 1L of PBS. All samples were run in duplicate with 7 controls on each 96-well plate consisting of a PVA/PVP/casein dilution buffer blank, 2 negative controls, 1 internal control and 3 positive controls. For each control serum/antigen pairing, % coefficient of variation (%CV) was calculated. If CV values for 3 or more of any control/antigen pairing on any plate exceeded 15%, the run was deemed unacceptable. Aliquots of individual bead suspensions were placed into 5mL of assay buffer per run in a volume sufficient to have 2500 beads of each classification in each well and wrapped in foil to protect the microbeads from photobleaching. Wells of a MultiScreen™ 96-well filter-bottom assay plate (Millipore) were pre-wet with 100uL of PBS/0.05% Tween-20

(TwPBS), which was removed with vacuum filtration. After vortexing, 50uL of bead suspension was added to each well and assay buffer was removed through vacuum filtration. Wells were washed twice with TwPBS, and TwPBS was removed through vacuum filtration after each wash. Serum dilutions were centrifuged at 16,000 x g for 10 minutes and 140 uL of each serum sample were pre-aliquoted into a V-bottom 96-well plate. 50uL of each serum sample was then aliquoted into its respective well on the filter-bottom plate containing the pre-aliquoted beads. The plate was covered and placed on a shaker for 1.5 hours at room temperature under aluminum foil at 700 rpm. After the incubation, the serum dilutions were removed through vacuum filtration and each well was washed with 100uL TwPBS, again using vacuum filtration. Secondary antibody stock solution was created by aliquoting 10uL of biotinylated mouse anti-human IgG (Southern Biotech) and 8uL of biotinylated mouse anti-human IgG_4 (Southern Biotech) in 5mL of assay buffer for a final mass of 50ng of IgG and 40ng of IgG_4 per well. The plate was covered, placed under aluminum foil and incubated at room temperature at 700 rpm for 45 minutes. After the incubation, buffer was removed from the wells through vacuum filtration and washed 3 times with TwPBS as before. 25uL of streptavidin-phycoerythrin (SAPE) (Molecular Probes) was added to 5mL of assay buffer for a final mass of 250ng of SAPE per well, and 50uL of this SAPE solution was added to each well. The plate was again covered, placed under aluminum foil and incubated at room temperature at 700 rpm for 30 minutes. After the incubation, the plate was again washed 3 times with TwPBS through vacuum filtration. To remove unbound antibody and to reduce background, 50uL of assay buffer was added to each well and the plate was incubated under aluminum foil at room temperature at 700 rpm for another 30 minutes. After this final incubation, the plate was washed one time with TwPBS through vacuum

filtration. 100uL of PBS at pH 7.2 was added to each well, covered and shaken for 30 seconds. Immediately afterwards, the plate was acquired on the Luminex machine using BioPlex software at 100 beads per classification.

Cut-off values. Antibodies were quantified by median fluorescence intensity, and total antibody response was characterized by the difference between the MFI of the sample and the negative control, or the background (MFI-bg). Cut-off values for this procedure have only been established for *Toxoplasma* SAG2, *Giardia* spp. VSP3 and VSP5, *Cryptosporidium* spp. Cp17, Cp23 and Cp2, and *E. histolytica* LecA. To establish the cut-off value for *Toxoplasma* SAG2 antigen, a 2.5 to 20 IU/mL standard curve was used to determine MFI values relative to known SAG2 antibody concentrations. The cut-off MFI value from this standard curve for SAG2 antibody was 314 units (Priest et al., 2015 (unpublished data)). For the two *Giardia* VSPs, 81 adults known to be seronegative were used and the highest 5% of values were dropped. Then, the remaining antibody responses were used to establish a cut-off at the mean plus 3 SD. The cut-off values for VSP3 and VSP5 were 358 and 233, respectively (Priest et al., 2015 (unpublished data)). For *Cryptosporidium* CpP2, Cp17 and Cp23, cut-off values were based on a receiver operating characteristic (ROC) curve based on Western blot data. For the Cp17 and Cp23 ROC curves, sera were obtained from the same 81 U.S. adults that were used for to obtain the VSP cut-off values. For Cp17, 44 adults were blot positive and 37 were blot negative; for Cp23, 60 were blot positive and 21 were blot negative. For CpP2, ROC curves were based off of sera obtained from 26 children in Haiti, among whom 15 were blot positive and 11 were blot negative. The cut-off values for CpP2, Cp17 and Cp23 were 89, 259 and 662, respectively (Priest et al., 2015 (unpublished data)). Notably, despite the

relatively low cut-off MFI value for CpP2, the ROC curve indicated that this cut-off point had 100% specificity and 100% sensitivity for this sample of Haitian children (Priest et al., 2015 (unpublished data)). Finally, for LecA, 65 American adults with no history of foreign travel were used. The highest three responses were eliminated and the cut-off was established as the mean + 3SD. The cut-off value for LecA was 302, which exhibited 100% sensitivity from ameabiasis patients (Moss et al., 2014) . To establish cut-offs for positive values for the seroconversion analysis of the remaining antigens, the distribution of MFI values for each antigen was analyzed using available data. MFI values for ETEC LT β subunit and *V. cholerae* toxin β subunit were bimodally distributed and negatively skewed, so a cut-off was established at the mean to avoid misclassification of seropositive children. Median MFI values were used as the cut-off for norovirus envelope proteins, and *Campylobacter* p18 and p39 antigens, with the understanding that some misclassification may occur immediately around the median. Due to uniform apparent negative values below the 90th percentile for *Salmonella* Group B & D LPS, the MFI value at the 90th percentile was used to enhance specificity. For ETEC LT β subunit, *V. cholerae* toxin β subunit, norovirus envelope proteins, and *Campylobacter* p18 and p39, and *Salmonella* Group B & D LPS, results were interpreted as the likelihood of a serological response above these assigned cut-off values, which are arbitrary relative to the established and tested cut-off MFI values for Cp17, Cp23, CpP2, LecA, VSP3 and VSP5 antibodies.

Imputation methods. To compensate for loss of children to follow-up, multiple imputation methods were used to impute missing follow-up seroprevalence data for all children who were enrolled in the study at baseline. The multiple imputation procedure

applied predictors of diarrhea identified during the baseline assessment (exclusive breastfeeding, water source, toilet type, socioeconomic status, feces on or around the toilet, shared toilet, gender and age) and all serological responses to adequately project co-occurring serological responses. An arbitrary, rather than monotone, missing pattern is anticipated, so missing values were imputed using a fully conditional specification (FCS) method that applies separate conditional distributions for each missing variable (Yuan, 2014). Imputed binary variables for serological responses were calculated based on predictors of MFI using multivariate linear regression. Covariates associated with seroprevalence at $\alpha \leq 0.35$ were included in the FCS imputation models for serological responses against each enteric pathogen. Data were imputed to create 25 total imputed datasets [127]. After modeling, parameter estimates were combined from the 25 imputed datasets to generate valid statistical inferences about the associations under study (Yuan, 2000). All analyses were run with both observed samples and imputed data, and analysis model covariates were derived directly from the models used for the regression imputations [129].

Statistical analysis. For all antibodies assessed in this study, the serological responses as measured by MFI were not normally distributed. Goodness-of-fit tests were performed on all log-transformed values for normal, lognormal, Weibull, gamma, beta and exponential distributions. If log-transformed variables fit a normal distribution, they were subject to parametric analyses. All variables that could not be transformed to fit a common statistical distribution were only analyzed for serological response relative to their cut-off values.

Prevalence of seroconversions against enteropathogens in this study were calculated using available paired data, and prevalence of serological responses against enteropathogens were calculated using both available and imputed data. The change in log-transformed MFI between baseline and follow-up (ΔMFI-bg) was compared between intervention and control groups using a paired t-test with Zeger robust variance estimator [130] where parametric analyses were possible. MFI-bg data for each antigen were dichotomized above and below their respective cut-off points at baseline and follow-up. Binary seroprevalence estimates were calculated among children in households randomized to intervention households and compared to children in control households using log binomial models. These models were run for both observed and imputed data to assess bias resulting from missing data [127]. For the seroconversion analyses using complete observations only, a child was considered to have seroconverted against a particular antigen if their MFI-bg values at baseline were below the cut-off but above the cut-off at follow-up. Seroconversion prevalence was compared with both observed and imputed data between intervention and control groups using log binomial models to calculate the relative risk of both seroprevalence and seroconversion against any specific antigen among children in the intervention vs. control arm. For these analyses, log-binomial models were favored over robust Poisson models as log-binomial models generally result in less bias and lower standard errors [131]. Both unadjusted and adjusted analyses were performed for available and imputed datasets, and adjusted analyses accounted for child age at baseline and the time elapsed between baseline and follow-up. We assessed confounding by other demographic, water and sanitation factors, even though we would expect reasonable balance between study arms on household and environmental factors collected at baseline (Zambrano et al., 2016, data not published).

Dichotomized MFI-bg values representing seroprevalence of antibody responses against each antigen were assessed relative to diarrhea prevalence among all children, with data combined between baseline and follow-up. Survey respondents were asked to recall whether the child had experienced diarrhea as per the WHO case definition within the previous 7 days. The relative risk for diarrheal disease in the previous 7 days in seropositive vs. seronegative children was compared for each antigen of interest using log binomial models accounting for repeated child measurements between baseline and follow-up. Unadjusted and adjusted analyses were performed for both available and imputed datasets.

Age-specific MFI-bg values for antibody against each antigen were calculated and plotted alongside seven-day diarrhea prevalence. Follow-up values for children in the intervention arm were omitted to properly represent age-specific seroprevalence against each antigen in the absence of the water filter intervention. Age-specific ΔMFI-bg values reflecting the change in serologic response between baseline and follow-up were calculated for both intervention and control arms. All analyses were performed using SAS V9 (SAS Institute, Cary, NC USA) and SAS-callable SUDAAN V11.0.1 (RTI International, Research Triangle Park, NC USA).

Ethical approval. The study protocol, survey instruments blood draw procedures and informed consent were reviewed and approved by the Emory University Institutional Review Board (Ref # 73615), the London School of Hygiene and Tropical Medicine Research Ethics Committee (Ref # 7711), the Rwandan National Ethics Committee (Ref

1497) and the National Health Research Committee of Rwanda (Ref # NHRC/2014/PROT/0163). The transfer of blood samples was governed by a Material Transfer Agreement between Emory University and the Rwanda Biomedical Center, executed September 2015. The analysis protocol and and transfer of samples from Emory University to the CDC was cleared by the Institutional Review Board of the Centers for Disease Control and Prevention (CDC/NCEZID #110415JP).

Results

Samples analyzed. Out of the 251 children who met our age eligibility criteria for enrollment in this seroconversion sub-study, 189 children who were 6 to 12 months-old at baseline were ultimately enrolled. Children were not enrolled if their caregiver did not consent to the blood draw (17.13%), the child was not at home during the baseline assessment (5.58%) or if the child was too ill to participate (0.80%). Of these 189 children, 19.05% were lost to follow-up due to refusals (3.17%), unsuccessful draws (0.53%), unavailability of the child (10.58%) or child illness (4.76%). Among the 189 children enrolled at baseline, samples were deemed insufficient if spots were less than approximately 90% filled in 36.51% of collected samples in the laboratory. As a result of this and loss of study subjects at follow-up, only 97 paired baseline and follow-up samples were available for the seroconversion analysis, with 34 (35.05%) and 68 (64.95%) of paired samples drawn from children in intervention and control households, respectively. For the overall seroprevalence study not incorporating seroconversion between baseline and follow-up, 120 samples were available for analysis from children at baseline and 152 were available from children at follow-up. Among the 120 samples drawn at baseline, 42 (35.00%) and 78 (65.00%) were available from intervention and

control households, respectively. Among the 152 samples drawn during follow-up, 62 (40.79%) and 90 (59.21%) of samples were drawn from children in intervention and control households (Supplemental Tables 7.1 and 7.2).

Parametric serological analyses. MFI-bg values for all norovirus antigens, *Toxoplasma* SAG2, *Salmonella* LPS-B and LPS-D, all *Cryptosporidium* antigens and *Campylobacter* p18 were successfully log-transformed to fit a normal distribution (Supplemental Table 7.3). There was no apparent or statistical difference in median serologic response, as measured by ΔMFI-bg, between intervention and control groups when examining serological responses against antigens for norovirus (St. Cloud strain), *Salmonella*, *Campylobacter*, *Giardia*, *E. histolytica*, enterotoxigenic *E. coli* and *V. cholerae*. No child appeared to produce any serological response to *Cryptosporidium* CpP2 peptide or *Toxoplasma* SAG2 throughout the course of the study; therefore, these antigens were dropped from subsequent analyses. Children in the intervention group did have median ΔMFI-bg values that were roughly equivalent to that for children in the control group for the Norwalk strain of norovirus and 23.0% of that of children in the control group for the Sydney strain of norovirus, but these results were not significant (p=0.483) (Table 7.1). Responses against *Cryptosporidium* Cp17 were appeared to be reduced for children in the intervention group (median ΔMFI-bg=47) compared to the control group (median ΔMFI-bg=260), although these results were not significant (p=0.212). Median responses against *Cryptosporidium* Cp23 and *Cryptosporidium* CpP2 appeared to be roughly equivalent between the two intervention arms. Median responses against *Campylobacter* p39, *Giardia* VSP3, *Giardia* VSP5, *E. histolytica* LecA, ETEC labile toxin β subunit and cholera toxin β subunit could not be directly compared using parametric analyses;

therefore, responses to these antigens were only examined using binomial variables derived from threshold cut-offs.

Serological responses by age. Known cut-off values for seropositivity are available for *Giardia* VSP3 and VSP5 antigens, *Cryptosporidium* Cp17 and Cp23 antigens, and *E. histolytica* LecA antigen; therefore, age-specific serological responses against these antigens were considered in terms of seroprevalence. Antigens of specific pathogens were aggregated so seroprevalence (as a proxy for cumulative exposure) could be considered for each pathogen. At follow-up, only observations from children in control households were used in order to provide a true seroprevalence measure in this population without any interference from the intervention. Seroprevalence of *Giardia* and *E. histolytica* responses remained low through the first year of life. *Giardia* seroprevalence increased sharply after 12 months and remained elevated through 24 months, although antibody responses appear to decrease after 24 months (Figure 1). *E. histolytica* responses remained markedly low throughout the study period, with a slight increase in seroprevalence after 18 months of age. Children appear to be exposed to *Cryptosporidium* at younger ages, as approximately 20% of children 6-8 months-old at baseline were seropositive. Like *Giardia*, *Cryptosporidium* seroprevalence increased markedly after 12 months, and antibody production remained elevated among children 12-24 months throughout the study period (Figure 1).

Cut-off values for seropositivity are not known for the remaining pathogens; therefore, reactivity against these antigens was assessed directly through median fluorescence intensity with background subtracted (MFI-bg) (Table 7.1). Age specific median MFI-

bg values were obtained, and log(MFI) of antibodies against each pathogen were plotted against age group (Figure 2). Generally, antibody reactivity against all antigens increased with age. Antibody responses for Sydney and St. Cloud norovirus increased steadily throughout the study period after 9 months of age. Antibody responses against Norwalk decreased initially after 9 months but then increased steadily, peaking at 21 months. Children were reactive against both *Campylobacter* strains at baseline, but their responses appeared to increase markedly after 12 months of age. Antibody responses against the EtxB and CtxB increased sharply after 9 months, peaking at 12 months and remaining elevated through 24 months. Responses to *Salmonella* LPS-B and LPS-D were low among all children of all ages, with reactivity only appearing to occur in children at 24 months.

Impact of intervention on serologic responses. Analyses were performed using both imputed (Tables 7.2a and 7.2b) and complete-case data (Supplemental Tables 7.4a and 7.4b) to calculate relative risks of both seroprevalence at follow-up and seroconversion between baseline and follow-up for all enteric antigens in this study. Prior to assessing the relative risks of serological response and seroconversion between intervention and control groups, tetrachoric correlation coefficients were calculated to determine whether follow-up serological data could be analyzed in isolation from the baseline serological data. Baseline and follow-up measures for seroprevalence were significantly correlated for *Giardia* VSP3 and VSP5 antigens, *Cryptosporidium* Cp17, envelope proteins for Norwalk, Sydney and St. Cloud strains of norovirus, *Campylobacter* p18 and p39 antigens, ETEC LT β-subunit and *V. cholerae* toxin β subunit (Supplemental Table 7.5). Correlation between baseline and follow-up measures indicates the potential for residual

expression of antibody spanning both study rounds. Because of this, we opted to measure both 1) raw seroprevalence against all antigens at follow-up and 2) seroconversion, measured as seroprevalence at follow-up among children who were negative at baseline.

Relative risk estimates were comparable for imputed and observed data (Tables 7.2a and 7.2b; 7.4a and 7.4b). At follow-up, seroprevalence and seroconversion estimates against *Giardia* VSP3 and VSP5 antigens appeared to be elevated among children in intervention households, although these associations were not significant (VSP3 seroconversion aRR: 1.35, 95%CI: 0.88-2.07; VSP5 seroconversion aRR: 1.46, 95%CI: 0.93-2.29). Seroprevalence and seroconversion against *Cryptosporidium* Cp17 was markedly reduced by 30% among children in the intervention group (seroprevalence aRR: 0.70, 95%CI: 0.54-0.91; seroconversion aRR: 0.72, 95%CI: 0.52-1.00). Seroprevalence and seroconversion against *Cryptosporidium* Cp23 appeared to be reduced in the intervention group, although these relationships were not statistically significant. Given the opposing intervention-attributable effects on *Giardia* and *Cryptosporidium* infection, the models examining the impact of the intervention on *Giardia* seroprevalence were stratified by *Cryptosporidium* seroprevalence at follow-up in order to assess potential interaction between the two pathogens. When *Cryptosporidium* Cp17 serological response was present, the filter generally did not affect *Giardia* VSP3 (aRR: 1.15, 95%CI: 0.68-1.93) or VSP5 (aRR: 1.19, 95%CI: 0.70-2.02) seroprevalence at follow-up; however, when Cp17 serological response was absent, the data suggested that seroprevalence against *Giardia* VSP3 and VSP5 may nearly double among children in households assigned to the intervention group (VSP3 aRR: 1.73, 95%CI: 0.79-3.75; VSP5 aRR: 2.01, 95%CI: 0.84-4.83), although these results were not statistically significant given the relatively

few number of children who were seronegative for Cp17 at follow-up. These results, along with relative sample sizes for each stratum, are presented in Table 7.3. No significant effect on seroprevalence or seroconversion was noted against the Norwalk, Sydney or St. Cloud strains of norovirus, although a moderate, albeit non-significant, reduction was noted for the Sydney strain (aRR=0.85, 95%CI: 0.67-1.09). The intervention did not significantly affect seroprevalence or seroconversion of antibody responses against *E. histolytica* LecA antigen, *Campylobacter* p18 and p39 antigens and the LT β subunits of ETEC and *V. cholerae*.

Association between seroprevalence against enteropathogens and diarrhea.

Seven-day prevalence of diarrheal disease nearly doubled in children with positive serological responses against *Giardia* VSP3 (aRR: 1.92, 95%CI: 1.33-3.25). Additionally, children appeared to be at higher risk for diarrheal disease at follow-up if they were seropositive against *Giardia* VSP5 (aRR: 1.78, 95%CI: 0.93-3.42), although this relationship was not statistically significant. Diarrhea prevalence was also significantly associated with seropositivity against *Cryptosporidium* Cp17 (aRR: 1.97, 95%CI: 1.11-3.51) but not with *Cryptosporidium* Cp23. A nearly 90% reduction in diarrhea prevalence was associated with serological response to *Salmonella* D LPS CHAPS (aRR: 0.12, 95%CI: 0.02-0.94). Serological responses against *Campylobacter* p18 and p39, *E. histolytica* LecA, any norovirus envelope protein, EtxB and CtxB were not associated with diarrhea prevalence (Table 7.4).

Association between nutritional status and diarrhea and seroprevalence

While there was no significant association between diarrhea and nutritional status at either baseline or follow-up, 25.0% of children in this sub-study who were designated as malnourished at baseline were reported to have had diarrhea in the previous week, compared to 16.9% of children with normal MUAC measurements (p=0.22).

As only one child at follow-up in this sub-population had MUAC < 11.5cm, analyses examining the relationship between seroprevalence and malnutrition were performed only on the baseline data. There was no association between moderate (MUAC= 11.6-12.5 cm) or severe (MUAC ≤ 11.5 cm) malnutrition and serological response. The relationship between reaction to *Cryptosporidium* Cp23 and MUAC (≤ 12.5cm) did appear to approach significance, particularly after adjusting for age (aRR: 2.73, 95%CI: 0.91-8.20, p=0.074), but this association was not observed for serological responses against *Cryptosporidium* Cp17. Overall, among the children at baseline who expressed antibody to Cp23, 43.8% appeared to be at least moderately malnourished (MUAC ≤ 12.5 cm).

Discussion

While parametric analyses did not indicate any meaningful differences in antibody responses to the antigens in our study between intervention and control groups, differences were noticed after MFI values were dichotomized along their designated cut-off points. Decreased seroprevalence and seroconversion against *Cryptosporidium* Cp17 was noted through analyses of both imputed and observed data. Conversely, seroconversion against *Giardia* VSP3 and VSP5 and *Salmonella* LPS-D appeared be higher in children living in intervention households.

There is also an as-yet unexplored possibility of biological interactions between the enteropathogens included in these analyses; infection with one pathogen may either exacerbate or attenuate the risk of infection with another pathogen [132]. Such an interaction may be exemplified by the opposing effects that the intervention had on *Giardia* and *Cryptosporidium* seroconversion. Among children seropositive for *Cryptosporidium*, the filter does not affect *Giardia* seroprevalence; however, among children seronegative for *Cryptosporidium*, the filter may be associated with increased *Giardia* seroconversion. Given that *Cryptosporidium* seroconversion generally occurred at earlier ages than *Giardia* seroconversion (Figure 1), it is realistic to assume that *Cryptosporidium* infections preceded most *Giardia* infections in this study among children who were seropositive for both. Given that the intervention significantly reduced seroprevalence and seroconversion against *Cryptosporidium*, any possible attenuation of *Giardia* infection by a prior *Cryptosporidium* infection through some biological mechanism, such as immunomodulation, may also be lower in the intervention group. The direct biological relationship between *Giardia* and *Cryptosporidium*, if one is present, has not been explored in-depth and should be defined. This study indicates that more holistic approaches may need to be practiced when assessing the joint epidemiologic and pathogenic associations between organisms in co-infection scenarios. Overall, though, the fact that the filter appears to significantly decrease *Cryptosporidium* seroconversion between baseline and follow-up in our study is encouraging. In the Global Enteric Multicenter Study (GEMS), stool-positive *Cryptosporidium* infection in children with moderate-to-severe diarrhea who were 12-23 months-old (the same age range as

children at follow-up in this study) nearly tripled the risk of death between their enrollment and follow-up periods[133]. enteropathogens

Serological responses to both *Giardia* and *Cryptosporidium* were associated with diarrheal disease in the previous 7 days, lending further support to the potential utility of using serological assays as an objective method to evaluate the health impact of household water treatment interventions [19] and to supplement self-reported health outcome data. This initially appears to contradict results by Kotloff et al., which detected *Giardia* more frequently in stool samples of controls than diarrhea cases; however, *Giardia* oocysts are present only intermittently in stool, compromising sensitivity [134]. Additionally, the use of culture and PCR-based methods for *Giardia* detection are only useful to identify current and active episodes, and serological studies may be useful in not only detecting active giardiasis cases but in linking previous *Giardia* infection to other long-term health outcomes, such as intestinal enteropathy [135].

Too few serological responses against *Salmonella* antigens were present for further analysis, and *Salmonella* seroconversion generally only occurred among the very oldest children in this sub-study population (Figure 2). Incidence of diarrheal disease in children throughout Sub-Saharan Africa generally peaks between 6-11 months of age and wanes after 24 months [75], which was demonstrated in this population of children in the baseline phase of this study (Zambrano et al., 2016, data not published). From the age-specific serological response data, it appears that children are not becoming infected with *Salmonella* during this crucial age window when they are at highest risk for diarrheal disease.

Notably, neither ETEC, *V. cholerae*, *Campylobacter*, or norovirus were associated with diarrheal disease, which is likely due to the short duration of disease relative to the duration of antibody response. This may indicate a need for longitudinal follow-up in shorter intervals in a similar population of children to record diarrheal disease closer to the time of infection, given that these pathogens tend to lead to short-term acute cases of diarrhea. Diarrheal disease attributed to protozoa, particularly *Giardia* species, can lead to persistent infection and duodenal inflammation [136], which may explain why associations between serological evidence of previous infection is associated with one-week prevalence of diarrhea.

There were some limitations in the design of this study that may affect the interpretation of these results. It is rare for household water treatment interventions to be blinded in RCTs for practical purposes, and this lack of blinding can lead to substantial reporting bias in which usage of the intervention and the intervention's effect on diarrheal disease are exaggerated [137]. If serological assays are applied as objective measures of enteric infection, it is possible that when implemented correctly, they may provide a more accurate depiction of health status than reliance on self-reported diarrheal disease alone [19,137]; however, diarrheal disease outcomes may have been affected by reporting and courtesy bias in this study. Another notable source of potential bias is imperfect intervention uptake. Among children in intervention households, 23.6% were reported to have consumed unfiltered water in the previous 24 hours. Given the already limited sample size among children in intervention households, stratifying on exclusive consumption of filtered water would have substantially compromised study power, but

exposure to unfiltered water among children in intervention households could have the effect of biasing our results towards the null.

From a laboratory analysis perspective, firm cut-off points for seropositivity were not available for EtxB, CtxB, *Campylobacter* p18/p39, *Salmonella* LPS or norovirus antigens. Among the antigens for which cut-off values were not available, parametric analyses were possible on log-transformed MFI-bg values for norovirus (Norwalk, Sydney and St. Cloud envelope), *Salmonella* LPS B and D CHAPS and *Campylobacter* p18. After examining the distribution of MFI-bg values for each antigen lacking a known cut-off point, relatively arbitrary cut-off points were made at the median, mean of 90^{th} percentile, as appropriate, and made in favor of sensitivity rather than specificity. As a result, some children who were seronegative were likely misclassified as seropositive for these antigens, thereby biasing results towards the null. For these antigens, the terms "seroprevalence" and "seroconversion" should be interpreted as being above and below these set cut-off points, as values above the cut-off cannot be confirmed as truly seropositive. To further enhance this analysis and analyses of future studies, definitive cut-off points should be established by applying known seropositive or seronegative sera to an ROC curve analysis for each antigen [138]. It is worth noting that the antigens for which cut-off points were previously established yielded significant and informative results in this study, both for their association with the water filter intervention and with diarrheal disease.

Other issues lie with the availability of antigens for this assay. The EtxB and CtxB extracted for this analysis are homologous, thereby limiting our ability to independently

attribute serological response to either ETEC or *V. cholerae*. EtxB is generally more immunostimulatory than CtxB, which appears to be consistent with results from our study [139]; however, due to cross-reactivity between the two homologues, seroprevalence against either of the two pathogens cannot be definitively characterized. A systematic way of differentiating between the serological response to both pathogens is likely not possible; therefore, interpretations at this time should be made with regard to exposure to the toxin itself rather than the causal pathogens. Rotavirus is typically vitally important to consider as a cause of diarrhea in children under two years-old; rotavirus-attributable incidence of moderate-to-severe diarrhea among children in this age group far outpaced that of other pathogens in the GEMS study [140]. It should be noted, though, that among children eligible for enrollment in this study at baseline, only 4.1% of children had not initiated their course of rotavirus vaccines, while 81.0% of children had received all three doses of rotavirus vaccine, as confirmed by examining the vaccination cards of all children enrolled in the RCT. If bead-based rotavirus immunoassay methods are developed and incorporated into multiplex enteric pathogen protocols, the target antigens should be exclusive of the viral glycoproteins targeted by the vaccine.

Given the large proportion of baseline samples deemed insufficient in the laboratory, multiple imputation was necessary to increase study power while making valid statistical inferences. Twenty-five imputed datasets were created in order to synthesize parameter estimates of generalized estimating equation (GEE) modeling of all datasets to a single risk estimate. Generally, analysis models should be based closely off of models used to complete the imputation to limit the possibility of introducing bias; however, this limited

the degree to which extraneous household and environmental factors could be considered for confounding. Additionally, the relatively sample size in turn yielded low seropositive counts, limiting the number of potential confounders that could be included in the imputation and analysis models. As a result, other environmental sources of infection were not considered in this study, either as potential confounders or exposures of interest.

Finally, we cannot make temporal inferences regarding the timing of infection with respect to either diarrheal disease or receipt of the intervention; however, inferences for timing of seroconversion with regard to the intervention are generally better, as households typically received the intervention within a couple of weeks of the initial baseline visit. Longitudinal follow-up with shorted follow-up rounds could have provided the opportunity to capture diarrheal disease closer to the time of infection, which may have provided richer data regarding pathogens that are associated with acute episodes of diarrhea, such as norovirus and *Campylobacter*. Frequent longitudinal collection may also provide a more thorough assessment of age-specific prevalence of individual pathogens.

Conclusion

This study suggests that serological testing of pathogen-specific antibodies can provide both measures of WASH intervention effectiveness and markers of diarrheal disease. The potential for between-pathogen interactions is present; as the water filter significantly reduced *Cryptosporidium* seroprevalence, it appeared to concurrently increase seroprevalence against *Giardia* antigens. Diarrheal disease appeared to be associated with protozoan pathogens only. Acute infections caused by other pathogens on the panel

may cause diarrhea, but this association may not be detected using these methods without frequent sampling intervals. Longitudinal intervention studies involving larger populations and repeated sampling would provide richer data that would enable further assessment of the utility of these serological approaches to evaluating household water treatment interventions.

Table 7.1. Comparison of median MFI values with background subtracted (MFI-bg), and the change in MFI-bg (ΔMFI-bg) from baseline to follow-up compared between children in intervention (LFS) and control households with a paired t-test.

		Overall	LFS	Control		
Antigen	Median (range) (MFI-bg)[1]	Median (range) ΔMFI-bg	Median (range) ΔMFI-bg	Median (range) ΔMFI-bg	β estimate	t (p-value)
Norovirus strain Norwalk	119 (-1 – 27964)	6 (-24337– 27959)	5 (-24337 – 27959)	7 (-23185 – 27168)	0.426	0.74 (0.459)
Norovirus strain Sydney	220 (-1 – 4734)	123 (-2143 – 4579)	42 (-686 –4435)	183 (-2143 – 4579)	0.360	-0.70 (0.483)
Norovirus strain St. Cloud	22 (-9 – 4380)	12 (-678 – 1645)	17 (-678 – 1645)	9 (-370 – 583)	0.446	1.13 (0.260)
Toxoplasma SAG2	4 (-5 – 27848)	1 (-38 – 107)	1 (-38 – 107)	1 (-37 – 62)	0.344	1.70 (0.091)
Salmonella LPS-B	5 (-2 – 6600)	1 (-58 – 6599)	4 (-58 – 6599)	0 (-33 – 4668)	0.336	1.29 (0.200)
Salmonella LPS-D	4 (-3 – 2121)	0 (-34-2118)	1 (-19 – 706)	0 (-34 – 2118)	0.374	1.56 (0.121)
Cryptosporidium CpP2(100)	9 (2 – 91)	1 (-67 – 57)	1 (-67 – 25)	1 (-31 – 57)	0.089	0.43 (0.670)
Cryptosporidium Cp17	118 (3 – 28866)	130 (-28547 – 28071)	47 (-995 – 28071)	260 (-28547 – 27611)	-0.768	-1.25 (0.212)
Cryptosporidium Cp23	113 (3 – 28862)	142 (-25072 – 27732)	142 (-1122 – 27732)	145 (-25072 – 27712)	0.159	0.26 (0.796)
Campylobacter p18	293 (6 – 27691)	57 (-24604 – 25095)	276 (-24571 – 25095)	16 (-24604 – 24157)	0.736	1.23 (0.222)
Campylobacter p39	77 (4 – 29051)	14 (-25283 – 26722)	17 (-24923 – 25615)	14 (-25283 – 26722)	NA	NA
Giardia VSP3	7 (-3 – 26343)	3 (-15915 – 24408)	5 (-24 – 24281)	2 (-15915 – 24408)	NA	NA
Giardia VSP5	6 (-3 – 26680)	2 (-22895 – 25624)	3 (-20293 – 24608)	2 (-22895 – 25624)	NA	NA
E. histolytica LecA	7 (-5 – 28833)	-1 (-1176 – 28818)	0 (-137 – 28818)	-2 (-1176 – 22769)	NA	NA
ETEC EtxB	23847 (9 – 29837)	5279 (-28941 – 29797)	9025 (-28941 – 29797)	2844 (-28179-29446)	NA	NA
Cholera CtxB	3381 (6 – 28565)	5423 (-24902 – 28539)	8395 (-23549 – 27674)	3829 (-24902 – 28539)	NA	NA

[1]Mean fluorescence intensity values can range from 1 to 32766 without background subtracted, but MFI values with background values subtracted (MFI-bg) can be negative [27].

Table 7.2a: Crude and adjusted risk ratios comparing Round 2 seroprevalence among children in the intervention and control groups who were 6-12 months-old at enrollment, Western Province, Rwanda June-September 2015. These data incorporate imputed values for samples deemed insufficient at the time of analysis. Adjusted analyses account for age and time between study rounds.

Antigen[1]	Cut-off (MFI-bg)	Method to Establish Cut-Off	Intervention (n=75) Crude Seroprevalence	Control (n=114) Crude Seroprevalence	Crude RR (95%CI, p-value)	Adjusted RR (95%CI, p-value)
Giardia						
VSP3	358	Mean + 3SD	31.2 (0.4160)	37.8 (0.3319)	1.25 (0.83-1.89, 0.275)	1.27 (0.85-1.90, 0.248)
VSP5	233	Mean + 3SD	30.8 (0.4101)	35.0 (0.3070)	1.34 (0.87-2.04, 0.179)	1.36 (0.83-1.89, 0.275)
Cryptosporidium						
Cp17	259	ROC	38.2 (0.5093)	81.8 (0.7175)	0.71 (0.54-0.92, 0.011)	0.70 (0.54-0.91, 0.007)
Cp23	662	ROC	36.0 (0.4805)	66.4 (0.5821)	0.82 (0.60-1.13, 0.228)	0.84 (0.62-1.16, 0.289)
E. histolytica						
LecA	302	Mean + 3SD	7.3 (0.0971)	8.6 (0.0758)	1.28 (0.48-3.45, 0.624)	1.16 (0.46-2.93, 0.760)
Salmonella						
LPS-B	38	90th percentile	19.2 (0.2565)	23.4 (0.2049)	1.25 (0.70-2.23, 0.449)	1.28 (0.71-2.28, 0.413)
LPS-D	13	90th percentile	17.9 (0.2379)	13.7 (0.1200)	1.98 (0.99-3.97, 0.053)	2.12 (1.05-4.29, 0.037)
Norovirus						
Norwalk	84	Median	45.7 (0.6091)	75.0 (0.6579)	0.93 (0.73-1.18, 0.530)	0.92 (0.73-1.17, 0.515)
Sydney	156	Median	45.2 (0.6021)	83.2 (0.7302)	0.82 (0.65-1.04, 0.107)	0.85 (0.67-1.09, 0.210)
St. Cloud	19	Median	45.0 (0.6000)	68.0 (0.5965)	1.01 (0.79-1.28, 0.962)	1.02 (0.81-1.29, 0.851)
Campylobacter						
p18	276	Median	53.1 (0.7077)	73.7 (0.6467)	1.09 (0.88-1.37, 0.423)	1.07 (0.87-1.33, 0.516)
p39	74	Median	54.6 (0.7280)	79.3 (0.6954)	1.05 (0.86-1.27, 0.645)	1.04 (0.86-1.25, 0.713)
Enterotoxigenic *E. coli*						
EtxB	15474	Mean	63.4 (0.8448)	91.1 (0.7989)	1.06 (0.92-1.22, 0.447)	1.01 (0.88-1.17, 0.863)
V. cholerae						
CtxB	9882	Mean	49.8 (0.6640)	74.8 (0.6558)	1.01 (0.80-1.28, 0.918)	0.98 (0.79-1.21, 0.863)

[1]No children were seropositive for *Toxoplasma* SAG2 or *Cryptosporidium* CpP2 antibody at baseline or follow-up; therefore, these analyses are not included.

Table 7.2b: Crude and adjusted risk ratios of imputed data comparing Round 2 seroprevalence among children in the intervention and control groups who were 6-12 months-old and seronegative at enrollment.

Antigen[1]	Cut-off (MFI-bg)	Method to Establish Cut-Off	Intervention Seroconversion n (%)	Control Seroconversion n (%)	Crude RR (95%CI, p-value)	Adjusted RR (95%CI, p-value)
Giardia						
VSP3	358	Mean + 3SD	30.2/73.9 (0.4080)	32.7/107.1 (0.3056)	1.34 (0.87-2.05, 0.184)	1.35 (0.88-2.07, 0.168)
VSP5	233	Mean + 3SD	29.7/73.7 (0.4026)	30.1/107.3 (0.2804)	1.44 (0.92-2.26, 0.114)	1.46 (0.93-2.29, 0.101)
Cryptosporidium						
Cp17	259	ROC	30.2/59.9 (0.5033)	63.2/92.6 (0.6828)	0.74 (0.53-1.02, 0.067)	0.72 (0.52-1.00, 0.050)
Cp23	662	ROC	31.2/68.4 (0.4553)	54.4/95.2 (0.5708)	0.80 (0.56-1.14, 0.210)	0.81 (0.58-1.14, 0.235)
E. histolytica						
LecA	302	Mean + 3SD	6.3/74.0 (0.0849)	8.6/113.0 (0.0765)	1.11 (0.39-3.13, 0.848)	0.95 (0.47-1.92, 0.890)
Salmonella						
LPS-B	38	90th percentile	18.2/73.0 (0.2499)	23.4/114.0 (0.2049)	1.22 (0.67-2.19, 0.514)	1.23 (0.68-2.23, 0.488)
LPS-D	13	90th percentile	15.8/69.0 (0.2290)	13.7/110.0 (0.1244)	1.84 (0.90-3.78, 0.096)	2.02 (0.97-4.22, 0.060)
Norovirus						
Norwalk	84	Median	23.9/42.8 (0.5585)	36.7/63.3 (0.5803)	0.96 (0.65-1.41, 0.844)	0.95 (0.64-1.41, 0.810)
Sydney	156	Median	22.4/43.4 (0.5157)	40.5/63.5 (0.6373)	0.81 (0.54-1.20, 0.286)	0.86 (0.58-1.29, 0.475)
St. Cloud	19	Median	35.0/63.0 (0.5556)	47.0/83.0 (0.5663)	0.98 (0.73-1.31, 0.897)	0.98 (0.74-1.30, 0.883)
Campylobacter						
p18	276	Median	31.4/48.64 (0.6464)	39.8/66.8 (0.5955)	1.08 (0.77-1.53, 0.643)	1.04 (0.71-1.52, 0.856)
p39	74	Median	35.3/51.2 (0.6898)	48.6/74.2 (0.6552)	1.05 (0.79-1.40, 0.718)	1.08 (0.77-1.53, 0.643)
Enterotoxigenic *E. coli*						
EtxB	15474	Mean	42.0/52.0 (0.8078)	63.4/85.3 (0.7487)	1.08 (0.88-1.32, 0.461)	1.02 (0.85-1.22, 0.859)
V. cholerae						
CtxB	9882	Mean	35.8/59.2 (0.6043)	57.9/95.4 (0.6069)	1.00 (0.74-1.34, 0.977)	0.97 (0.73-1.28, 0.807)

Table 7.3: Association between seroconversion against *Giardia* VSP3 and VSP5 antigens and intervention status, stratified by *Cryptosporidium* Cp17 serological response.

	Antigen[1]	Intervention Seroconversion n (%)	Control Seroconversion n (%)	Crude RR (95%CI, p-value)	Adjusted RR (95%CI, p-value)
Cryptosporidium Cp17 response present	VSP3	15.4/38.4 (0.4017)	29.7/81.7 (0.3634)	1.10 (0.65-1.87, 0.711)	1.15 (0.68-1.93, 0.711)
	VSP5	15.7/38.4 (0.4079)	29.1/81.7 (0.3565)	1.15 (0.68-1.94, 0.615)	1.19 (0.70-2.02, 0.525)
Cryptosporidium Cp17 response absent	VSP3	15.4/36.6 (0.4212)	7.9/32.3 (0.2438)	1.74 (0.80-3.79, 0.165)	1.73 (0.79-3.75, 0.168)
	VSP5	15.4/36.6 (0.4223)	6.8/32.3 (0.2116)	2.01 (0.84-4.83, 0.119)	2.01 (0.84-4.83, 0.117)

Table 7.4: Association between serological response and seven-day diarrhea prevalence. All adjusted models are adjusted for age and socioeconomic status.

		Diarrhea prevalence		
Antigen	Serologic response present (%)	Serologic response absent (%)	Unadjusted RR (95% CI, p-value)	Adjusted RR (95%CI, p-value)
Giardia				
VSP3	17/73 (0.2329)	36/222 (0.1622)	1.44 (0.85-2.44, 0.179)	1.56 (0.83-2.94, 0.168)
VSP5	16/71 (0.2254)	37/224 (0.1652)	1.36 (0.79-2.34, 0.258)	1.46 (0.77-2.76, 0.246)
Cryptosporidium				
Cp17	31/142 (0.2183)	22/153 (0.1438)	1.52 (0.94-2.47, 0.091)	1.78 (1.02-3.12, 0.044)
Cp23	23/118 (0.1949)	30/177 (0.1695)	1.15 (0.69-1.90, 0.586)	1.20 (0.69-2.11, 0.518)
E. histolytica				
LecA	4/15 (0.2667)	49/280 (0.1750)	1.52 (0.65-3.59, 0.335)	1.30 (0.56-2.99, 0.539)
Salmonella				
LPS-B	3/38 (0.0789)	50/257 (0.1946)	0.41 (0.14-1.21, 0.105)	0.44 (0.15-1.29, 0.136)
LPS-D	1/37 (0.0270)	52/258 (0.2016)	0.13 (0.02-0.94, 0.043)	0.14 (0.02-0.92, 0.041)
Norovirus				
Norwalk	30/167 (0.1796)	23/128 (0.1797)	1.00 (0.59-1.69, 0.999)	1.03 (0.61-1.74, 0.908)
Sydney	30/168 (0.1786)	23/127 (0.1811)	0.99 (0.60-1.61, 0.955)	1.08 (0.64-1.83, 0.771)
St. Cloud	33/177 (0.1864)	20/121 (0.1653)	1.13 (0.69-1.83, 0.627)	1.20 (0.73-1.99, 0.474)
Campylobacter				
p18	31/167 (0.1856)	22/128 (0.1719)	1.08 (0.63-1.84, 0.777)	1.04 (0.59-1.83, 0.895)
p39	33/164 (0.2012)	20/131 (0.1527)	1.32 (0.73-2.38, 0.353)	1.32 (0.75-2.32, 0.330)
Enterotoxigenic *E. coli*				
EtxB	34/177 (0.1921)	19/118 (0.1610)	1.19 (0.68-2.09, 0.537)	1.39 (0.64-3.04, 0.402)
V. cholerae				
CtxB	28/139 (0.2014)	25/156 (0.1603)	1.26 (0.73-2.17, 0.410)	1.34 (0.68-2.66, 0.399)

[1]Adjusted for age and socioeconomic status

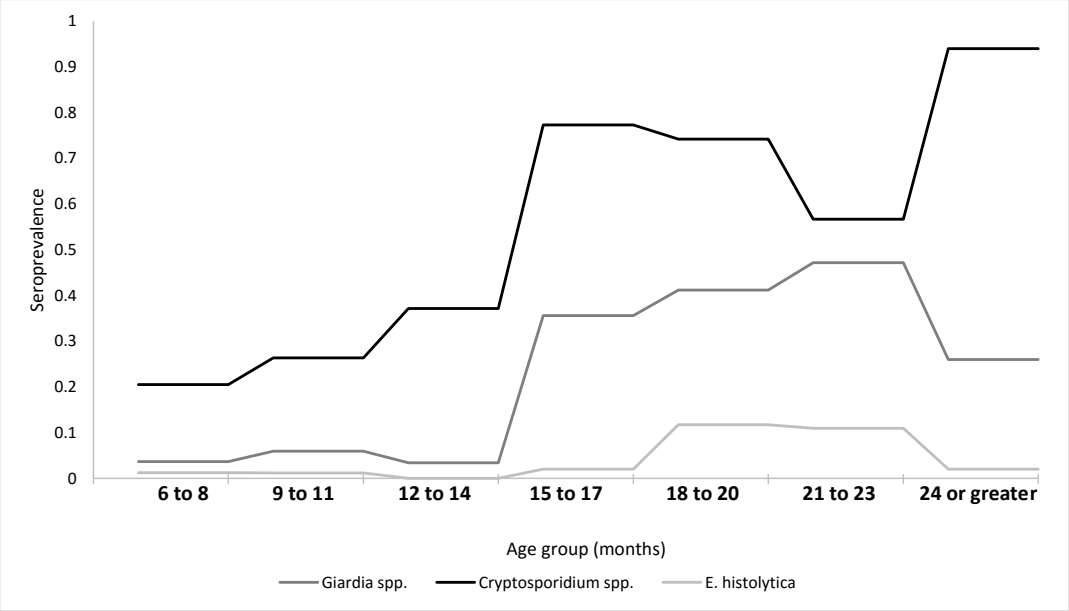


Figure 7.1. Age-specific prevalence of serological responses by age group among children 6 to 24 months-old against *Giardia* spp. VSP3 and VSP5 antigens, *Cryptosporidium* spp. Cp17 and Cp23 antigens and *E. histolytica* LecA antigen, based off of known MFI cut-off values.

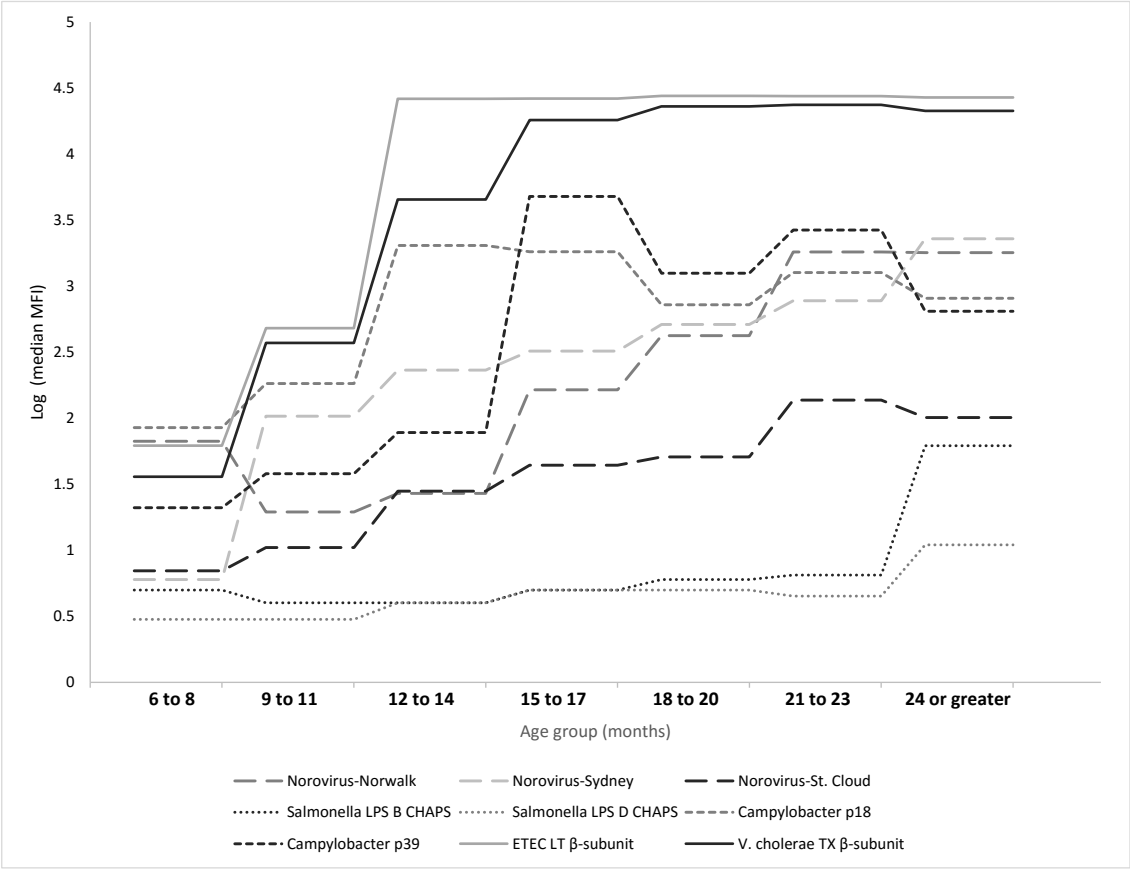


Figure 7.2. Age-specific median MFI values reflecting antibody responses against norovirus envelope proteins (Norwalk, St. Cloud and Sydney strains), *Salmonella* LPS B and D groups, *Campylobacter* p18 and p39, EtxB and CtxB.

SUPPLEMENTAL MATERIAL

Supplemental Table 7.1: Categorization of enrollment and sample loss issues at baseline.

Study Arm	Total number of children approached	Blood draw not successful	Caregiver refused	Child not at home	Child too sick for blood draw	Total samples collected
Intervention	111	1 (0.90%)	24 (21.62%)	10 (9.01%)	1 (0.90%)	75 (67.5%)
Control	140	2 (1.43%)	19 (13.57%)	4 (2.86%)	1 (0.71%)	114 (81.43%)
Total	251	3 (1.20%)	43 (17.13%)	13 (5.58%)	2 (0.80)	189 (75.30%)
Assessment of Sample Acceptability for Laboratory Analysis (among samples collected)						
		N	Sample Insufficient		Samples available for analysis	
Intervention		75	33 (44.00%)		42 (56.00)	
Control		114	36 (31.58%)		78 (68.42%)	
Total		189	69 (36.51%)		120 (63.49)	

Supplemental Table 7.2: Categorization of enrollment and sample loss issues at Round 2.

Study Arm	Total number of children approached	Blood draw not successful	Caregiver refused	Child not at home	Child too sick for blood draw	Total samples collected
Intervention	75	0	1 (1.33%)	8 (10.67%)	4 (5.33%)	62 (82.67%)
Control	114	1 (0.88%)	5 (4.39%)	12 (10.53%)	5 (4.39%)	91 (79.82%)
Total	189	1 (0.53%)	6 (3.17%)	20 (10.58%)	9 (4.76%)	153 (80.95%)
Assessment of Sample Acceptability for Laboratory Analysis (among samples collected)						
		N	Sample Insufficient		Samples available for analysis	
Intervention		62	0		62 (100.00%)	
Control		91	1 (1.10%)		90 (98.90%)	
Total		153	1 (0.65%)		152 (99.35%)	

Supplemental Table 7.3: Goodness-of-fit tests and distribution characteristics of all log-transformed antibody-specific median fluorescence intensity values, centered on log(MFI)=8.5.

Antigen	GOF test	Statistic (p-value)	Distribution	Characteristics	Further transformation needed?
Norovirus strain Norwalk	Kolmogorov-Smirnov	D=0.057 (0.134)	Normal	Mean=10.16, SD=3.45	No
Norovirus strain Sydney	Kolmogorov-Smirnov	D=0.058 (0.123)	Gamma	Scale=sigma (0.566); Shape=alpha (19.07) Mean=10.79, SD=2.47	Q-Q plot indicates that distribution approximates normal
Norovirus strain St. Cloud	Kolmogorov-Smirnov	D=0.037 (>0.150)	Normal	Mean=10.32, SD=1.96	No
Toxoplasma SAG2	Kolmogorov-Smirnov	D=0.060 (0.091)	Normal	Mean=9.09. SD=1.15	No
Salmonella LPS B CHAPS	Kolmogorov-Smirnov	D=0.062 (0.078)	Gamma	Scale=sigma (0.296); Shape=alpha (32.07) Mean=9.50, SD=1.68	Q-Q plot indicates that distribution approximates normal
Salmonella LPS D CHAPS	Cramer-von Mises	W^2=0.110 (0.180)	Lognormal	Scale=zeta (2.188); Shape=sigma (0.125) Mean=8.99, SD=1.13	No
CpP2(100) peptide	Kolmogorov-Smirnov	D=0.066 (0.044)	~Normal	Mean=8.74, SD=0.75	No
Cp17	Kolmogorov-Smirnov	D=0.060 (0.092)	Normal	Mean=11.18. SD=3.02	No
Cp23	Cramer-von Mises	W^2=0.115 (0.067)	Weibull	Scale=sigma (12.561); Shape=C (4.20) Mean=11.42, SD=3.06	No
Campylobacter p18	Kolmogorov-Smirnov	D=0.035 (>0.150)	Normal	Mean=10.31, SD=2.70	No
Giardia VSP3	Native and log-transformed values do not yield a known distribution. Non-parametric analysis required.				
Giardia VSP5					
ETEC labile TX Beta subunit					
Cholera TX Beta subunit					
Campylobacter p39					
E. histolytica LecA					

Supplemental Table 7.4a: Complete-case analysis of crude and adjusted risk ratios comparing Round 2 seroprevalence among children in the intervention (LFS) and control groups who were 6-12 months-old at enrollment. These data incorporate available values only; samples deemed insufficient at the time of analysis were not included.

Antigen	Cut-off (MFI-bg)	Method to Establish Cut-Off	Intervention (n=62) Crude Seroprevalence	Control (n=90) Crude Seroprevalence	Crude RR (95%CI, p-value)	Adjusted RR (95%CI, p-value)
Giardia						
VSP3	358	Mean + 3SD	26 (0.4194)	30 (0.3333)	1.26 (0.83-1.90, 0.277)	1.29 (0.86-1.94, 0.222)
VSP5	233	Mean + 3SD	26 (0.4194)	28 (0.3111)	1.35 (0.88-2.06, 0.168)	1.38 (0.91-2.11, 0.132)
Cryptosporidium						
Cp17	259	ROC	32 (0.5161)	64 (0.7111)	0.72 (0.55-0.95, 0.021)	0.70 (0.53-0.92, 0.010)
Cp23	662	ROC	30 (0.4839)	53 (0.5889)	0.81 (0.60-1.11, 0.193)	0.80 (0.59-1.08, 0.142)
E. histolytica						
LecA	302	Mean + 3SD	6 (0.0968)	7 (0.0778)	1.24 (0.42-3.70, 0.695)	0.93 (0.29-3.04, 0.906)
Salmonella						
LPS-B	38	90th percentile	15 (0.2419)	17 (0.1889)	1.32 (0.70-2.46, 0.390)	1.39 (0.75-2.60, 0.298)
LPS-D	13	90th percentile	15 (0.2419)	11 (0.1222)	1.91 (0.94-3.88, 0.072)	1.96 (0.96-3.96, 0.063)
Norovirus						
Norwalk	84	Median	39 (0.6290)	60 (0.6667)	0.94 (0.74-1.20, 0.634)	0.95 (0.75-1.20, 0.681)
Sydney	156	Median	37 (0.5968)	66 (0.7333)	0.78 (0.61-1.00, 0.052)	0.89 (0.74-1.07, 0.207)
St. Cloud	19	Median	45 (0.7258)	68 (0.7556)	0.97 (0.76-1.25, 0.830)	0.98 (0.77-1.25, 0.884)
Campylobacter						
p18	276	Median	44 (0.7097)	59 (0.6556)	1.08 (0.87-1.35, 0.477)	1.07 (0.86-1.34, 0.557)
p39	74	Median	45 (0.7258)	63 (0.7000)	1.02 (0.83-1.26, 0.825)	1.02 (0.84-1.23, 0.868)
Enterotoxigenic *E. coli*						
EtxB	15474	Mean	53 (0.8548)	72 (0.8)	1.06 (0.92-1.23, 0.415)	1.01 (0.86-1.20, 0.874)
V. cholerae						
CtxB	9882	Mean	42 (0.6774)	59 (0.6556)	1.02 (0.81-1.28, 0.890)	0.98 (0.80-1.20, 0.873)

Supplemental Table 7.4b: Complete-case analysis of crude and adjusted risk ratios comparing Round 2 seroprevalence among children in the intervention (LFS) and control groups who were 6-12 months-old and seronegative at enrollment.

Antigen[1]	Cut-off (MFI-bg)	Method to Establish Cut-Off	Intervention Crude Seroprevalence	Control Crude Seroprevalence	Crude RR (95%CI, p-value)	Adjusted RR (95%CI, p-value)
Giardia						
VSP3	358	Mean + 3SD	11/33 (0.3333)	15/59 (0.2542)	1.31 (0.68-2.51, 0.415)	1.31 (0.68-2.51, 0.415)
VSP5	233	Mean + 3SD	11/33 (0.3333)	14/59 (0.2373)	1.40 (0.72-2.73, 0.316)	1.40 (0.72-2.73, 0.318)
Cryptosporidium						
Cp17	259	ROC	15/27 (0.5556)	34/53 (0.6415)	0.85 (0.57-1.26, 0.422)	0.81 (0.53-1.23, 0.304)
Cp23	662	ROC	14/32 (0.4375)	29/52 (0.5577)	0.77 (0.48-1.21, 0.257)	0.80 (0.51-1.25, 0.320)
E. histolytica						
LecA	302	Mean + 3SD	4/33 (0.1212)	5/62 (0.0806)	1.50 (0.43-5.22, 0.52)	0.94 (0.22-4.10, 0.936)[2]
Salmonella						
LPS-B	38	90th percentile	9/32 (0.2813)	13/63 (0.2063)	1.41 (0.66-2.98, 0.373)	1.49 (0.70-3.16, 0.303)
LPS-D	13	90th percentile	8/29 (0.2759)	8/59 (0.1356)	1.97 (0.82-4.70, 0.129)	1.99 (0.83-4.77, 0.123)
Norovirus						
Norwalk	84	Median	9/21 (0.4286)	20/35 (0.5714)	0.75 (0.42-1.33, 0.324)	0.77 (0.45-1.31, 0.341)
Sydney	156	Median	7/22 (0.3182)	23/38 (0.6053)	0.53 (0.28-1.02, 0.058)	0.71 (0.40-1.25, 0.237)
St. Cloud	19	Median	15/30 (0.5000)	27/47 (0.5745)	0.84 (0.53-1.31, 0.436)	0.83 (0.53-1.29, 0.403)
Campylobacter						
p18	276	Median	15/26 (0.5769)	21/36 (0.5833)	0.98 (0.64-1.51, 0.930)	0.94 (0.59-1.50, 0.810)
p39	74	Median	16/26 (0.6154)	28/42 (0.6667)	0.90 (0.61-1.32, 0.592)	0.93 (0.56-1.54, 0.787)
Enterotoxigenic *E. coli*						
EtxB	15474	Mean	18/24 (0.7500)	31/45 (0.6889)	1.08 (0.79-1.46, 0.640)	1.03 (0.64-1.65, 0.896)
V. cholerae						
CtxB	9882	Mean	15/28 (0.5357)	29/52 (0.5577)	0.94 (0.62-1.43, 0.764)	0.94 (0.63-1.42, 0.779)

[1]No children were seropositive for *Toxoplasma* SAG2 or *Cryptosporidium* CpP2 antibody at baseline or follow-up; therefore, these analyses are not included.

[2]Due to lack of model convergence, RR calculated through Poisson rather than log binomial model.

Supplemental Table 7.5: Baseline and Round 2 follow-up antigen-specific seroprevalence using available acceptable DBS samples (n=120) and tetrachoric correlation coefficients (ρ) and likelihood ratio(LR) χ^2 test to assess the correlation of baseline and Round 2 follow-up seroprevalence estimates.

Antigen	Baseline seroprevalence	Follow-up seroprevalence	ρ +/- SE	LR χ^2 (p-value)
VSP3	7 (5.83%)	56 (36.84%)	0.9869+/-0	12.00 (0.001)
VSP5	7 (5.83%)	54 (35.53%)	0.9880+/-0	12.36 (<0.001)
Cp17	23 (19.17%)	96 (63.16%)	0.4561+/-0.1774	5.25 (0.022)
Cp27	16 (13.33%)	83 (54.61%)	0.3600+/-0.1880	3.19 (0.074)
LecA	2 (1.67%)	13 (8.55%)	0.5254+/-0.3146	2.06 (0.151)
B LPS CHAPS	2 (1.67%)	32 (21.05%)	0.2971+/-0.3494	0.66 (0.415)
D LPS CHAPS	10 (8.33%)	26 (17.11%)	0.0734+/-0.2519	0.09 (0.771)
Norwalk	50 (41.67%)	99 (65.13%)	0.5154+/-0.1348	10.63 (0.001)
Sydney	46 (38.33%)	103 (67.76%)	0.7618+/-0.0988	24.28 (<0.001)
St. Cloud	43 (35.83%)	113 (59.79%)	0.2445+/-0.1243	3.63 (0.057)
Cp18	43 (35.83%)	103 (67.76%)	0.4860+/-0.1449	8.51 (0.004)
Cp39	36 (30.00%)	108 (71.05%)	0.4815+/-0.1575	7.14 (0.008)
ETEC LT β subunit	31 (25.83%)	125 (82.24%)	0.6147+/-0.1590	9.65 (0.002)
Cholera LT β	20 (16.67%)	101 (66.45%)	0.9856+/-0	17.85 (0.001)

Supplemental Table 7.6: Polychoric correlation coefficients and associated likelihood ratio χ^2 tests to assess correlation between serological responses against all pathogens on the panel.

Variable	With Variable	ρ	SE	LR χ^2	p-value	Variable	With Variable	ρ	SE	LR χ^2	p-value
Norwalk	**Sydney**	0.46435	0.06841	35.9865	<.0001	**VSP3**	**Camp. Cp18**	0.30857	0.08846	10.9869	0.0009
Norwalk	**St. Cloud**	0.55789	0.06308	53.2631	<.0001	**VSP3**	**Camp. Cp39**	0.36384	0.08659	15.2749	<.0001
Norwalk	**VSP3**	0.11037	0.09336	1.3796	0.2402	**VSP3**	**E. histo. LecAP**	0.36733	0.13225	6.7561	0.0093
Norwalk	**VSP5**	0.07282	0.09140	0.6313	0.4269	**VSP5**	**Cp17**	0.35218	0.08358	15.5698	<.0001
Norwalk	**Cp17**	0.18940	0.07956	5.4617	0.0194	**VSP5**	**Cp23**	0.36288	0.08395	16.2969	<.0001
Norwalk	**Cp23**	0.21231	0.08095	6.5609	0.0104	**VSP5**	**SALM. B LPS**	0.21728	0.10920	3.8003	0.0512
Norwalk	**SALM. B LPS**	0.15608	0.10281	2.2471	0.1339	**VSP5**	**Salm. D LPSP**	-0.04295	0.12142	0.1249	0.7238
Norwalk	**Salm. D LPSP**	0.10119	0.10610	0.8999	0.3428	**VSP5**	**ETEC**	0.45067	0.08050	24.8512	<.0001
Norwalk	**ETEC**	0.27814	0.07667	12.1359	0.0005	**VSP5**	**V. cholerae**	0.32557	0.08476	13.2060	0.0003
Norwalk	**V. cholerae**	0.19690	0.07935	5.9157	0.0150	**VSP5**	**Camp. Cp18**	0.30700	0.08620	11.4639	0.0007
Norwalk	**Camp. Cp18**	0.19379	0.07881	5.8182	0.0159	**VSP5**	**Camp. Cp39**	0.37934	0.08356	17.5825	<.0001
Norwalk	**Camp. Cp39**	0.23208	0.07793	8.3891	0.0038	**VSP5**	**E. histo. LecAP**	0.45467	0.12183	11.2157	0.0008
Norwalk	**E. histo. LecAP**	0.23175	0.13705	2.7187	0.0992	**Cp17**	**Cp23**	0.74296	0.04736	105.9978	<.0001
Sydney	**St. Cloud**	0.52907	0.06535	46.7495	<.0001	**Cp17**	**SALM. B LPS**	0.29950	0.09746	8.6159	0.0033
Sydney	**VSP3**	0.20137	0.09224	4.5642	0.0326	**Cp17**	**Salm. D LPSP**	0.17340	0.10428	2.6864	0.1012
Sydney	**VSP5**	0.24885	0.08873	7.3615	0.0067	**Cp17**	**ETEC**	0.52155	0.06567	45.5163	<.0001
Sydney	**Cp17**	0.28599	0.07716	12.5947	0.0004	**Cp17**	**V. cholerae**	0.47127	0.06854	36.6426	<.0001
Sydney	**Cp23**	0.29137	0.07897	12.4274	0.0004	**Cp17**	**Camp. Cp18**	0.41803	0.07131	28.2530	<.0001
Sydney	**SALM. B LPS**	0.12445	0.10369	1.4173	0.2338	**Cp17**	**Camp. Cp39**	0.49552	0.06711	40.8357	<.0001
Sydney	**Salm. D LPSP**	0.06997	0.10669	0.4278	0.5131	**Cp17**	**E. histo. LecAP**	0.23177	0.13380	2.8497	0.0914
Sydney	**ETEC**	0.33704	0.07468	18.0374	<.0001	**Cp23**	**SALM. B LPS**	0.16788	0.10338	2.5705	0.1089
Sydney	**V. cholerae**	0.22599	0.07887	7.7828	0.0053	**Cp23**	**Salm. D LPSP**	0.04903	0.10883	0.2026	0.6526
Sydney	**Camp. Cp18**	0.09201	0.08067	1.2899	0.2561	**Cp23**	**ETEC**	0.51141	0.06835	40.9663	<.0001
Sydney	**Camp. Cp39**	0.17914	0.07936	4.9311	0.0264	**Cp23**	**V. cholerae**	0.31815	0.07756	15.1289	0.0001
Sydney	**E. histo. LecAP**	0.13487	0.13912	0.9241	0.3364	**Cp23**	**Camp. Cp18**	0.38181	0.07494	22.1000	<.0001
St. Cloud	**VSP3**	0.34524	0.08604	14.2033	0.0002	**Cp23**	**Camp. Cp39**	0.40616	0.07396	25.0749	<.0001
St. Cloud	**VSP5**	0.20550	0.08906	5.1070	0.0238	**Cp23**	**E. histo. LecAP**	0.12978	0.13869	0.8622	0.3531
St. Cloud	**Cp17**	0.33560	0.07535	17.5944	<.0001	**SALM. B LPS**	**Salm. D LPSP**	0.58091	0.09022	28.0675	<.0001
St. Cloud	**Cp23**	0.42043	0.07293	27.3326	<.0001	**SALM. B LPS**	**ETEC**	0.44774	0.09383	18.2748	<.0001
St. Cloud	**SALM. B LPS**	0.24490	0.09960	5.6962	0.0170	**SALM. B LPS**	**V. cholerae**	0.45035	0.08959	20.1834	<.0001
St. Cloud	**Salm. D LPSP**	0.41417	0.09407	16.1048	<.0001	**SALM. B LPS**	**Camp. Cp18**	0.14278	0.10279	1.8891	0.1693
St. Cloud	**ETEC**	0.38387	0.07335	23.2776	<.0001	**SALM. B LPS**	**Camp. Cp39**	0.39503	0.09558	14.4286	0.0001
St. Cloud	**V. cholerae**	0.32854	0.07560	16.8389	<.0001	**SALM. B LPS**	**E. histo. LecAP**	0.19263	0.15954	1.4064	0.2357
St. Cloud	**Camp. Cp18**	0.29147	0.07672	13.1906	0.0003	**Salm. D LPSP**	**ETEC**	0.18475	0.10530	2.9709	0.0848
St. Cloud	**Camp. Cp39**	0.30796	0.07625	14.7398	0.0001	**Salm. D LPSP**	**V. cholerae**	0.23693	0.10232	5.0712	0.0243
St. Cloud	**E. histo. LecAP**	0.30438	0.13068	4.9533	0.0260	**Salm. D LPSP**	**Camp. Cp18**	0.05102	0.10627	0.2298	0.6316
VSP3	**VSP5**	0.97968	0.00968	241.1020	<.0001	**Salm. D LPSP**	**Camp. Cp39**	0.10119	0.10610	0.8999	0.3428
VSP3	**Cp17**	0.38298	0.08419	17.6760	<.0001	**Salm. D LPSP**	**E. histo. LecAP**	0.22002	0.15985	1.8080	0.1787
VSP3	**Cp23**	0.37827	0.08513	16.9967	<.0001	**ETEC**	**V. cholerae**	0.85023	0.03365	160.3988	<.0001
VSP3	**SALM. B LPS**	0.18891	0.11292	2.7162	0.0993	**ETEC**	**Camp. Cp18**	0.36906	0.07314	21.9547	<.0001
VSP3	**Salm. D LPSP**	0.05188	0.12153	0.1819	0.6697	**ETEC**	**Camp. Cp39**	0.57903	0.06080	59.5838	<.0001
VSP3	**ETEC**	0.47081	0.08184	25.6590	<.0001	**ETEC**	**E. histo. LecAP**	0.46432	0.13298	10.0340	0.0015
VSP3	**V. cholerae**	0.30821	0.08764	11.2091	0.0008	**V. cholerae**	**Camp. Cp18**	0.29717	0.07639	13.7776	0.0002
Camp. Cp18	**Camp. Cp39**	0.78105	0.04151	130.3752	<.0001	**V. cholerae**	**Camp. Cp39**	0.44054	0.07022	31.5721	<.0001
Camp. Cp18	**E. histo. LecAP**	0.10616	0.13824	0.5835	0.4450	**V. cholerae**	**E. histo. LecAP**	0.29512	0.13125	4.6433	0.0312
Camp. Cp39	**E. histo. LecAP**	0.30726	0.13547	4.7083	0.0300						

Supplemental Table 7.7: Adjusted models used to calculate 1) the association between intervention status and Round 2 seroprevalence (SP) against any particular antigen. All adjusted models measuring the association between seroprevalence (across both rounds) and diarrheal disease were adjusted by age and socioeconomic status and are not depicted in this table.

Adjusted log-binomial models examining effect of intervention status on Round 2 seroprevalence	
Outcome	**Variables Included in Adjusted model**
Cp17 R2-SP	Intervention arm, shared sanitation, age and breastfeeding status at baseline
Cp23 R2-SP	Intervention arm, shared sanitation, water source, breastfeeding status at baseline
VR2-SP3 R2-SP	Intervention arm, gender
VR2-SP5 R2-SP	Intervention arm, gender
Norwalk R2-SP	Intervention arm, shared sanitation
Sydney R2-SP	Intervention arm, toilet area cleanliness
St. Cloud R2-SP	Intervention arm, shared sanitation
LecA R2-SP	Intervention arm, age
LPS-B R2-SP	Intervention arm, shared sanitation
LPS-D R2-SP	Intervention arm, socioeconomic status and age
Cp18 R2-SP	Intervention arm, gender, water source, shared sanitation
Cp39 R2-SP	Intervention arm, age, toilet area cleanliness, water source and socioeconomic status
EtxB R2-SP	Intervention arm, breastfeeding status at baseline, water source, toilet area cleanliness, socioeconomic status and age
CtxB R2-SP	Intervention arm, breastfeeding status at baseline, age and gender

Chapter 8 – Conclusion

The first two aims of this dissertation, covered in Chapters 5 and 6, utilized the cross-sectional data collected during the baseline phase of this cluster-randomized controlled trial in Western Province, Rwanda. The first study sought to characterize the associations between various household water and sanitation factors, water quality and reported one-week prevalence of diarrhea among children under 5 years-old. The second study aimed to describe the associations between various stove characteristics and cooking behaviors, household air pollution and upper and lower acute respiratory infections among children under 5 years-old. The third aim sought to examine the use of serological assays for enteropathogen exposure as an objective disease marker to supplement diarrheal disease outcome data. Antibody responses against various pathogens were assessed relative to 1) intervention arm classification at baseline and the second round of follow-up and 2) reported childhood diarrhea in the previous seven days.

For the first two aims, a multi-stage logistic regression modeling approach was used to characterize the associations between various water, sanitation and hygiene factors and cooking factors with diarrheal disease and respiratory infection, respectively. Administrative sectors were randomized at a 3:1 ratio into intervention and control groups, and villages were selected at a 1:1 ratio using population proportional selection. Within villages, up to 10 households were selected, and all children under 5 years-old in each household were enrolled in the study. This multi-stage approach accounted for clustering at the village and the household level using a Taylor series linearization approach with a between-cluster variance estimator [141]. Given that this modeling

procedure typically yields odds ratios, which are inappropriate for common outcomes, model-adjusted predicted marginal one-week prevalence estimates of diarrhea or ALRI, as appropriate, were estimated in order to calculate prevalence ratios [71,72]. Generally, logistic regression with predicted margins is preferred for complex survey data analysis. Logistic regression avoids model convergence issues common with log-binomial and modified Poisson models and prevents variance overestimation issues presented by Robust Poisson methods [131]. For the third aim, analyses were performed on both observations for which complete data were available and observations with complete and imputed data. Multiple imputation was performed to account for baseline and follow-up sample loss in order to provide valid statistical inferences of modeled parameter estimates [126]. Missing values were imputed using fully conditional specification methods [142] using variables that were loosely associated with the variable for which values were being imputed ($p<0.4$). This procedure yielded 25 imputed datasets, which were condensed using PROC MIANALYZE to yield valid parameter estimates. All subsequent analyst model covariates were assessed for confounding using the imputation model as the full model, and the results of imputed analyses were compared with analyses using observed data only to assess bias. Relative risks of seroconversion with regard to intervention status, and the relative risk of seropositivity with regard to reported diarrheal disease, were estimated through log binomial modeling. Survey methods, such as those applied for the first two aims, were not applied for this study, given the small average cluster size for this sub-study (1.49 children per village) and the complexities of accounting for stratification, clustering and weighting during the imputation procedure itself [143].

The first analysis found that sanitation status significantly affected diarrhea prevalence. Toilet-area cleanliness, indicated by the presence of feces around the toilet, and shared toilet facilities were significant risk factors for diarrhea after controlling for water source type, household proximity to water sources, toilet type, age, gender, household crowding and socioeconomic status. This study lends further evidence to the body of literature that suggests that shared sanitation increases the risk of diarrheal disease [78,79]. Composting toilets nearly doubled the risk of diarrhea, which may be due to inadequate waste management practices [82]. The majority of this diarrheal burden was actually attributed to households with non-shared vs. shared composting toilets. Given the substantial investment that installation of composting toilets requires, stakeholders should be made aware of the continuing importance of sanitation awareness, outreach and training. While only nine households practiced open defecation, this practice appeared to substantially increase diarrhea risk, even while it was associated with low fecal contamination of water. Specific water source types also appear to significantly affect diarrheal disease and are also associated with poor drinking water quality. Surface water sources appear to elevate risk while "improved" water sources, such as standpipes and hand pumps or boreholes significantly increased the risk of disease. Follow-up analyses should incorporate census data or an urbanization index before ascribing excess risk to the water sources themselves, as these water sources may merely be a proxy for household density. Notably, piped water sources and both protected and unprotected dug wells appeared to be highly protective.

While the second study found that only socioeconomic status was associated with ALRI, several factors appeared to be associated with personal $PM_{2.5}$ exposure. Among children,

rondereza stoves appeared to be significantly protective, which may be due to their configuration. *Rondereza* stoves are generally stationary, higher off the ground and attached to a wall, and children may not gather around these stoves as they do with three-stone fires. Cooking within the sitting room led to significantly more exposure compared to cooking in a separate kitchen for both primary cooks and children, which supports the implementer's messaging that cooking should occur outside of the primary living spaces. Wood and charcoal use was significantly protective among children compared to wood use only. Kerosene lamp usage is associated with lower exposure among primary cooks but higher exposure among children, indicating a need to explore what behavioral factors are linked to both kerosene lamp usage and $PM_{2.5}$ exposure. An interquartile increase in $PM_{2.5}$ exposure in this population nearly triples the risk of ALRI, which aligns with a recent RCT that elucidated the exposure-response relationship between personal carbon monoxide exposures linked to biomass burning and ALRI [94]. Diarrhea is significantly correlated with both ALRI and cough or cold, and ALRI prevalence is inversely correlated with MUAC. This aligns with previous research that has that ALRI incidence increases among children with recent diarrheal disease [111] and that both diarrhea and MUAC are associated with ALRI [109].

Results from the third study indicated that the water filter intervention significantly decreased seroconversion against *Cryptosporidium*, and by extension, protected children from *Cryptosporidium* infection. Contemporaneously, however, the data suggested that the water filter may increase *Giardia* seroconversion. Further stratification of this association by the presence or absence of a serological response against *Cryptosporidium* indicated that the risk of *Giardia* seroconversion might be exacerbated in the absence of

a serological response against *Cryptosporidium*. The potential biological interaction between these two enteric protozoa should be explored in more depth, as *Cryptosporidium* pathogenesis may affect host factors that mediate susceptibility to giardiasis, such as structural and cellular characteristics of the intestinal epithelium [144] which may affect *Giardia* trophozoite attachment. Immunomodulation may also play a role in susceptibility to giardiasis through activation of innate and humoral immune responses [145,146]. While this explanation is purely hypothetical at this point, other interactions between protozoa have been described [132] and researchers are beginning to examine the roles of the intestinal microbiome, innate immunity and host-parasite interactions in *Giardia* susceptibility [147,148]. *Giardia* and *Cryptosporidium* point seroprevalence are both positively associated with seven-day period prevalence of diarrhea, indicating that serological responses against these protozoa can be a useful objective measure to supplement self-reported diarrhea outcomes. The relationships between serological responses against the other enteropathogens in this study and intervention status and diarrhea can be bolstered by defining firm cut-off points for seropositivity. Given that diarrhea attributed to protozoa is experienced more chronically relative to diarrhea attributed to viral and bacterial causes, the association between serological responses to other enteropathogens in this study should be further explored with longitudinal analyses incorporating shorter follow-up intervals. Future longitudinal analyses can also characterize the timing of *Giardia* seroconversion with regard to *Cryptosporidium* exposure, which is vital in considering potential biological interactions between these two protozoa.

Strengths of this overall study include the size of the study population and representativeness of the source population. The sector-level randomization approach appeared effective at balancing intervention and control arms along all measured potential confounders, which will ease interpretation of the results of the overall CRT. Random selection of villages and households yielded basic demographic characteristics in our study population that aligned closely with data collected from larger representative surveys, such as the 2010 Rwandan Demographic and Health Survey [12] and the 2012 Population and Housing Census [51,61]. Paired with the fact that the statistical approaches used in the cross-sectional studies of the first two aims accounted for differential probabilities of selection and village-level clustering, the results from these studies can be extrapolated and generalized to the whole population of Western Province. Household characteristics, such as toilet type, stove type, stove number and cooking location, were directly observed by our enumerators. Water sources were ascertained during the survey using pictures presented to the respondents. Objective exposure and disease metrics, such as personal PM2.5 exposure monitoring, COHb, SPO_2 and blood pressure measurement, and serological markers of previous enteropathogen infection bolstered the otherwise subjective methods used in household surveys, which are subject to recall and courtesy biases. The multiple imputation procedures used for the seroconversion study indicated that little bias was contributed by the missing data points.

These studies were limited by a couple of important factors. Household selection was based off of the Rwandan Government's 2012 *Ubudehe* list. Given that the study began in 2014, these lists did not include people who had recently moved into the village since the list was released, who may differ in important but unmeasured ways from those who

had resided in the village for a longer period of time. These lists also included households that had since moved out of our study villages, which reduced the total number of truly eligible households available for selection into our study. Together, these issues resulted in varying cluster sizes, as we were not able to enroll our target of 10 households per village. Temporally linking disease outcomes with exposures was problematic throughout the survey. Water quality, measured by TTC counts, was not associated with diarrhea, which may be due to the fact that water quality was ascertained after reported diarrhea cases began. While ALRI was associated with $PM_{2.5}$ exposure, ALRI was ascertained on the first survey day while $PM_{2.5}$ exposure was measured over the ensuing 48 hours; therefore, an assumption of continuous exposure was made for the purpose of analyzing these results. Longitudinal sampling for serological data may help directly ascertain when seroconversions against specific pathogens are occurring and to develop hypotheses related to interactions between pathogens. Serological studies to develop firm MFI cut-off points for antibody reactions against all pathogens in this study will be necessary to further explore the impact of this water filter intervention more broadly on enteric pathogen seroconversion.

Overall, this dissertation outlines household risk factors for both diarrheal disease and pneumonia in young children. Some of the results were surprising and may warrant secondary analyses to confirm. For example, standpipes and boreholes, which were installed to enable community access to improved water sources, appear to significantly increase diarrheal disease relative to protected springs. Urbanization and household density should be incorporated to further characterize this relationship. Non-shared household-level composting toilets, intended to improve both household-level sanitation,

are also associated with increased diarrheal disease, which may indicate issues with household-level waste management and hygiene. Given the highly protective effect of piped water and dug wells, these interventions should be favored to improve water sources. Dangerous levels of $PM_{2.5}$ exposure were observed broadly across all exposure categories. While this cross-sectional study was conducted before the intervention was distributed, this study prevents evidence in favor of reducing concurrent stove use and of moving household cooking activities away from primary living areas. Finally, seroconversion against *Giardia* and *Cryptosporidium* antigens may be differentially affected by the water filter intervention, and both serological responses appear to be associated with reported diarrheal disease. Further refinement of cut-off points for other antigens will allow associations to be explored. This dissertation supports the use of serological markers to objectively ascertain both exposure and disease. The findings of this dissertation should assist organizations working on water, sanitation and household air pollution research and identified associations and potential interactions to be further explored by researchers.

References

1. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet [Internet]. 2012;380:2197–223. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23245608

2. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet [Internet]. 2012;380:2224–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23245609

3. WHO. Air Quality Guidelines: Global Update, 2005. Geneva, Switzerland; 2005.

4. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet [Internet]. 2012;380:2095–128. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L366291525

5. Barstow CK, Ngabo F, Rosa G, Majorin F, Boisson S, Clasen T, et al. Designing and Piloting a Program to Provide Water Filters and Improved Cookstoves in Rwanda. 2014;9:1–12.

6. Rosa G, Majorin F, Boisson S, Barstow C, Johnson M, Kirby M, et al. Assessing the impact of water filters and improved cook stoves on drinking water quality and household air pollution: a randomised controlled trial in Rwanda. PLoS One [Internet]. 2014;9:e91011. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3948730&tool=pmcentrez

&rendertype=abstract

7. Martin WJ, Glass RI, Araj H, Balbus J, Collins FS, Curtis S, et al. Household air pollution in low- and middle-income countries: health risks and research priorities. PLoS Med. [Internet]. 2013;10:e1001455. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3672215&tool=pmcentrez &rendertype=abstract

8. Rylance J, Gordon SB, Naeher LP, Patel A, Balmes JR, Adetona O, et al. Household air pollution: a call for studies into biomarkers of exposure and predictors of respiratory disease. Am. J. Physiol. Lung Cell. Mol. Physiol. [Internet]. 2013;304:L571–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23457186

9. MAL-ED Network Investigators. The MAL-ED Study: A Multinational and Multidisciplinary Approach to Understand the Relationship Between Enteric Pathogens, Malnutrition, Gut Physiology, Physical Growth, Cognitive Development, and Immune Responses in Infants and Children Up to 2 Years of. Clin. Infect. Dis. [Internet]. 2014;59 Suppl 4:S193–206. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25305287

10. Black RE, Allen LH, Bhutta Z a, Caulfield LE, de Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet [Internet]. 2008];371:243–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18207566

11. Guerrant RL, DeBoer MD, Moore SR, Scharf RJ, Lima A a M. The impoverished gut--a triple burden of diarrhoea, stunting and chronic disease. Nat. Rev. Gastroenterol. Hepatol. [Internet]. Nature Publishing Group; 2013;10:220–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23229327

12. National Institute of Statistics Rwanda. 2010 Demographic and Health Survey, Rwanda. 2010.

13. “Rwanda Bureau of Standards.” Potable Water — Specification. ICS 13.060.20 Rwanda; 2011.

14. Bartram J, Pedley S. MICROBIOLOGICAL ANALYSES. In: Bertram J, Ballance R, editors. Water Qual. Monit. A Pract. Guid. to Des. Implement. Freshw. Qual. Stud. Monit. Program. United Nations Environment Programme and the World Health Organization; 1996.

15. WHO. Guidelines for Drinking-Water Quality. Fourth Edi. Geneva, Switzerland: Gutenberg; 2011.

16. WHO. Integrated Management of Childhood Illness. Geneva, Switzerland; 2005.

17. Berkley J, Mwangi I, Griffiths K, Mithwani S, English M, Newton C. Assessment of Severe Malnutrition Among Hospitalized Children in Rural Kenya. 2014;294:591–7.

18. Rasmussen J, Andersen a, Fisker a B, Ravn H, Sodemann M, Rodrigues a, et al. Mid-upper-arm-circumference and mid-upper-arm circumference z-score: the best predictor of mortality? Eur. J. Clin. Nutr. [Internet]. Nature Publishing Group; 2012;66:998–1003. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22805497

19. Crump J a, Mendoza CE, Priest JW, Glass RI, Monroe SS, Dauphin L a, et al. Comparing serologic response against enteric pathogens with reported diarrhea to assess the impact of improved household drinking water quality. Am. J. Trop. Med. Hyg. [Internet]. 2007;77:136–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17620645

20. Lammie PJ, Moss DM, Brook Goodhew E, Hamlin K, Krolewiecki A, West SK, et al. Development of a new platform for neglected tropical disease surveillance. Int. J. Parasitol. [Internet]. Australian Society for Parasitology Inc.; 2012;42:797–800. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22846784

21. Brussow H, Sidoti J, Link H, Hoang Y, Barclay D, Dirren H, et al. Age-specific prevalence of antibody to enterotoxigenic Escherichia coli in Ecuadorian and German children. J. Infect. Dis. 1990;162:974–7.

22. Ungar B, Gilman R, Lanata C, Perez-Schael. Seroepidemiology of Cryptosporidium infection in two Latin American populations. J. Infect. Dis. 1988;157:551–6.

23. Vitral C, Yoshida C, Lemos E, Teixera C, Gaspar A. Age-specific prevalence of antibodies to hepatitis A in children and adolescents from Rio de Janeiro, Brazil, 1978 and 1995: relationship of prevalence to environmental factors. Mem. Inst. Oswaldo Cruz [Internet]. 1998;93:1–5. Available from: http://dx.doi.org/10.1590/S0074-02761998000100001

24. Khanna B, Cutler A, Israel N, Perry M, Lastovica A, Fields P, et al. Use caution with serologic testing for Helicobacter pylori infection in children. J. Infect. Dis. 1998;178:460–5.

25. Lindkvist P, Asrat D, Nilsson I. Age at acquisition of Helicobacter pylori infection: comparison of a high and a low prevalence country. Scand J Infect Dis. 1996;28:181–4.

26. Steinberg EB, Mendoza CE, Glass R, Arana B, Lopez MB, Mejia M, et al. PREVALENCE OF INFECTION WITH WATERBORNE PATHOGENS : A SEROEPIDEMIOLOGIC STUDY IN CHILDREN 6 – 36 MONTHS OLD IN SAN JUAN SACATEPEQUEZ , GUATEMALA. 2004;70:83–8.

27. Moss DM, Priest JW, Hamlin K, Derado G, Herbein J, Petri W a, et al. Longitudinal evaluation of enteric protozoa in haitian children by stool exam and multiplex serologic assay. Am. J. Trop. Med. Hyg. [Internet]. 2014 ;90:653–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24591430

28. Priest JW, Moss DM, Visvesvara GS, Jones CC, Li A, Isaac-Renton JL. Multiplex assay detection of immunoglobulin G antibodies that recognize Giardia intestinalis and

Cryptosporidium parvum antigens. Clin. Vaccine Immunol. [Internet]. 2010;17:1695–707. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2976096&tool=pmcentrez&rendertype=abstract

29. Moss DM, Priest JW, Hamlin K, Derado G, Herbein J, Petri W a, et al. Longitudinal evaluation of enteric protozoa in Haitian children by stool exam and multiplex serologic assay. Am. J. Trop. Med. Hyg. [Internet]. 2014;90:653–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24591430

30. Torres-Dosal A, Pérez-Maldonado IN, Jasso-Pineda Y, Martínez Salinas RI, Alegría-Torres J a, Díaz-Barriga F. Indoor air pollution in a Mexican indigenous community: evaluation of risk reduction program using biomarkers of exposure and effect. Sci. Total Environ. [Internet]. 2008;390:362–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18036639

31. Dherani M, Pope D, Mascararenhas M, Smith KR, Weber M, Bruce N. Indoor air pollution from unprocessed solid fuel use and pneumonia risk in children aged under five years: a systematic review and meta-analysis. Bull. World Health Organ. [Internet]. 2008;86:390–8. Available from: http://www.who.int/bulletin/volumes/86/5/07-044529.pdf

32. Smith KR, Bruce N, Balakrishnan K, Adair-Rohani H, Balmes J, Chafe Z, et al. Millions dead: how do we know and what does it mean? Methods used in the comparative risk assessment of household air pollution. Annu. Rev. Public Health [Internet]. 2014;35:185–206. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24641558

33. European Union Energy Initiative. Biomass Energy Strategy (BEST), Rwanda. Brussels, Belgium; 2009. Report No.: Project Number 01.2457.8-007.24.

34. Puumalainen T, Quiambao B, Abucejo-Ladesma E, Lupisan S, Heiskanen-Kosma

T, Ruutu P, et al. Clinical case review: a method to improve identification of true clinical and radiographic pneumonia in children meeting the World Health Organization definition for pneumonia. BMC Infect. Dis. [Internet]. 2008;8:95. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2492864&tool=pmcentrez&rendertype=abstract

35. Guarnieri MJ, Diaz J V, Basu C, Diaz A, Pope D, Smith KR, et al. Effects of woodsmoke exposure on airway inflammation in rural guatemalan women. PLoS One [Internet]. 2014;9:e88455. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3953023&tool=pmcentrez&rendertype=abstract

36. WHO. Workshop Resources: Indoor Air Pollution and Household Energy Monitoring. Geneva Switzerland; 2005.

37. Pennise D, Smith K, Naumoff K. Indoor Air Pollution Measurement Options. Berkeley, California;

38. WHO. Pollution levels and personal exposure. Eval. Househ. energy Heal. Interv. A Cat. methods. Geneva, Switzerland; 2009. p. 18–22.

39. Risom L, Moller P, Loft S. Oxidative stress-induced DNA damage by particulate air pollution. Mutat. Res. 2005;592:119–37.

40. Valavanidis a, Fiotakis K, Bakeas E, Vlahogianni T. Electron paramagnetic resonance study of the generation of reactive oxygen species catalysed by transition metals and quinoid redox cycling by inhalable ambient particulate matter. Redox Rep. [Internet]. 2005;10:37–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15829110

41. Riddick DS, Lee C, Ramji S, Chinje EC, Cowen RL, Williams KJ, et al. CANCER

CHEMOTHERAPY AND DRUG METABOLISM. 2005;33:1083–96.

42. Banerjee A, Mondal NK, Das D, Ray MR. Neutrophilic inflammatory response and oxidative stress in premenopausal women chronically exposed to indoor air pollution from biomass burning. Inflammation [Internet]. 2012;35:671–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21769440

43. Dutta A, Ray MR, Banerjee A. Systemic inflammatory changes and increased oxidative stress in rural Indian women cooking with biomass fuels. Toxicol. Appl. Pharmacol. [Internet]. Elsevier Inc.; 2012;261:255–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22521606

44. Park WY, Goodman RB, Steinberg KP, Ruzinski JT, Ii FR, Park DR, et al. Cytokine Balance in the Lungs of Patients with Acute Respiratory Distress Syndrome.

45. Schütte H, Lohmeyer J, Rosseau S, Ziegler S, Siebert C, Kielisch H, et al. Bronchoalveolar and systemic cytokine profiles in patients with ARDS, severe pneumonia and cardiogenic pulmonary oedema. Eur. Respir. J. [Internet]. 1996 [cited 2014 Nov 2];9:1858–67. Available from: http://erj.ersjournals.com/content/9/9/1858

46. Sack U, Scheibe R, Wötzel M, Hammerschmidt S, Kuhn H, Emmrich F, et al. Multiplex analysis of cytokines in exhaled breath condensate. Cytometry. A [Internet]. 2006 [cited 2014 Jan 23];69:169–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16496377

47. Kellum JA, Kong L, Fink MP, Weissfeld LA, Yealy DM, Pinksky MR, et al. Understanding the Inflammatory Cytokine Response in Pneumonia and Sepsis. Arch Intern Med. 2007;167:1655–63.

48. Sprague AH, Khalil R a. Inflammatory cytokines in vascular dysfunction and vascular disease. Biochem. Pharmacol. [Internet]. 2009;78:539–52. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2730638&tool=pmcentrez

&rendertype=abstract

49. Harrison DG, Guzik TJ, Lob HE, Madhur MS, Marvar PJ, Thabet R, et al. Inflammation , Immunity , and Hypertension. 2011;132–40.

50. Li J, Li J, He J, Nan J, Guo Y, Xiong C. Atorvastatin Decreases C-Reactive Protein-Induced Inflammatory Response in Pulmonary Artery Smooth Muscle Cells by Inhibiting Nuclear Factor- κ B Pathway. 2010;28:8–14.

51. NISR. 2012 Population and Housing Census, National Institute of Statistical Research, Rwanda.

52. Barstow CK, Ngabo F, Rosa G, Majorin F, Boisson S, Clasen T, et al. Designing and piloting a large-scale project to provide water filters and improved cookstoves in Rwanda. Data not Publ. 2013;

53. Rogers E. Diffusion of Innovations. 5th Editio. New York, New York: Free Press; 2003.

54. Becker M. The Health Belief Model and Personal Health Behavior. Health Educ. Monogr. 1974;2:324–473.

55. Thomas E, Barstow CK, Rosa G, Majorin F, Clasen T. Use of Remotely Reporting Electronic Sensors for Assessing Use of Water Filters and Cookstoves in Rwanda. Environ. Sci. Technol. 2013;47:13602–10.

56. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data- or tears: An application to educational enrollment in states of India. Demography. 2001;38:115–32.

57. Kolenikov S. The Use of Discrete Data in PCA : Theory , Simulations , and Applications to Socioeconomic Indices. 2004;

58. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan. [Internet]. 2006 [cited 2015 Jun 16];21:459–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17030551

59. U.S. Environmental Protection Agency. Water by Membrane Filtration Using Modified membrane-Thermotolerant Escherichia coli Agar (Modified mTEC). 2009.

60. WHO JMP & UNICEF. Progress on Sanitation and Drinking Water: 2015 Update and MDG Assessment. 2015.

61. National Institute of Statistics Rwanda. Fourth Population and Housing Census, Rwanda, 2012: Characteristics of households and housing. 2012.

62. Naghavi M, Wang H, Lozano R, Davis A, Liang X, Zhou M, et al. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet [Internet]. Elsevier Ltd; 2015;385:117–71. Available from: http://dx.doi.org/10.1016/S0140-6736(14)61682-2

63. Prüss-Ustün A, Bartram J, Clasen T, Colford JM, Cumming O, Curtis V, et al. Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: a retrospective analysis of data from 145 countries. Trop. Med. Int. Health [Internet]. 2014;19:894–905. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24779548

64. Clasen T, Pruss-Ustun A, Mathers CD, Cumming O, Cairncross S, Colford JM. Estimating the impact of unsafe water, sanitation and hygiene on the global burden of disease: evolving and alternative methods. Trop. Med. Int. Health [Internet]. 2014;19:884–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24909205

65. Onda K, LoBuglio J, Bartram J. Global access to safe water: accounting for water quality and the resulting impact on MDG progress. Int. J. Environ. Res. Public Health

[Internet]. 2012 [cited 2015 Sep 20];9:880–94. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3367284&tool=pmcentrez&rendertype=abstract

66. Bain R, Cronk R, Hossain R, Bonjour S, Onda K, Wright J, et al. Global assessment of exposure to faecal contamination through drinking water based on a systematic review. Trop. Med. Int. Health [Internet]. 2014 ;19:917–27. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4255778&tool=pmcentrez&rendertype=abstract

67. Wolf J, Prüss-Ustün A, Cumming O, Bartram J, Bonjour S, Cairncross S, et al. 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 [Internet]. 2014 ;19:928–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24811732

68. Clasen T, Boisson S. Assessing the Health Impact of Water Quality Interventions in Low-Income Settings: Concerns Associated with Blinded Trials and the Need for Objective Outcomes. Environ. Health Perspect. [Internet]. 2015; Available from: http://ehp.niehs.nih.gov/15-10532

69. WHO & UNICEF. Handbook: Integrated Management of Childhood Illness. 2005.

70. Rwandan Ministry of Health. Integrated Community Case Management of Childhood Illness, Ministry of Health, Rwanda: Trainer's Guide. 2014.

71. Graubard BI, Korn EL. Predictive Margins with Survey Data. 1999;55:652–9.

72. Bieler GS, Brown GG, Williams RL, Brogan DJ. Estimating model-adjusted risks, risk differences, and risk ratios from complex survey data. Am. J. Epidemiol. [Internet]. 2010;171:618–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20133516

73. Hosmer DW, Lemeshow S, Sturdivant R. Model-Building Strategies and Methods for Logistic Regression. 2013;

74. Fawzy A, Arpadi S, Kankasa C, Sinkala M, Mwiya M, Thea DM, et al. Early weaning increases diarrhea morbidity and mortality among uninfected children born to HIV-infected mothers in Zambia. J. Infect. Dis. [Internet]. 2011;203:1222–30. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3069726&tool=pmcentrez&rendertype=abstract

75. Fischer Walker CL, Perin J, Aryee MJ, Boschi-Pinto C, Black RE. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. BMC Public Health [Internet]. BioMed Central Ltd; 2012;12:220. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3323412&tool=pmcentrez&rendertype=abstract

76. Baber RM. MODELING SITE SUITABILITY FOR COMMUNITY WATER ACCESS POINTS IN THE MAYANGE SECTOR OF RWANDA. Redlands, CA; 2009.

77. Clasen T, Roberts I, Rabie T, Schmidt W, Cairncross S. Interventions to improve water quality for preventing diarrhoea (Review). Cochrane Collab. 2009;

78. Fuller J a, Clasen T, Heijnen M, Eisenberg JNS. Shared sanitation and the prevalence of diarrhea in young children: evidence from 51 countries, 2001-2011. Am. J. Trop. Med. Hyg. [Internet]. 2014;91:173–80. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4080558&tool=pmcentrez&rendertype=abstract

79. Heijnen M, Cumming O, Peletz R, Chan GKS, Brown J, Baker K, et al. Shared sanitation versus individual household latrines: A systematic review of health outcomes. PLoS One. 2014;9.

80. Pickering AJ, Julian TR, Marks SJ, Mattioli MC, Boehm AB, Schwab KJ, et al. Fecal contamination and diarrheal pathogens on surfaces and in soils among Tanzanian households with and without improved sanitation. Environ. Sci. Technol. [Internet]. 2012;46:5736–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22545817

81. Fuhrmeister ER, Schwab KJ, Julian TR. Estimates of Nitrogen, Phosphorus, Biochemical Oxygen Demand, and Fecal Coliforms Entering the Environment Due to Inadequate Sanitation Treatment Technologies in 108 Low and Middle Income Countries. Environ. Sci. Technol. [Internet]. 2015;49:11604–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26320879

82. Mukasine B, Country PK, Context HC, Policy W, National T, Iii E, et al. ECOSAN in Rwanda: A baseline study to identify challenges and opportunities. 2014.

83. Gundry S, Wright J, Conroy R. A systematic review of the health outcomes related to household water quality in developing countries. 2004;1–13.

84. Levy K. Does Poor Water Quality Cause Diarrheal Disease? Am. J. Trop. Med. Hyg. [Internet]. 2015;93:899–900. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26438028

85. Luby SP, Halder AK, Huda TM, Unicomb L, Islam MS, Arnold BF, et al. Microbiological Contamination of Drinking Water Associated with Subsequent Child Diarrhea. Am. J. Trop. Med. Hyg. [Internet]. 2015;93:904–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26438031

86. Vos T, Barber RM, Bell B, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet [Internet]. 2015;386:1990–2013. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673615606924

87. Rosenthal J. The real challenge for cookstoves and health: more evidence. Ecohealth [Internet]. Springer US; 2015;12:8–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25691140

88. You D, Hug L, Ejdemyr S, Idele P, Hogan D, Mathers C, et al. Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the un Inter-Agency Group for Child Mortality Estimation. Lancet [Internet]. World Health Organization. Published by Elsevier Ltd/Inc/BV. All rights reserved.; 2015;386:2275–86. Available from: http://dx.doi.org/10.1016/S0140-6736(15)00120-8

89. Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, Brauer M, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;386:2287–323.

90. Gordon SB, Bruce NG, Grigg J, Hibberd PL, Kurmi OP, Hubert Lam K, et al. Respiratory risks from household air pollution in low and middle income countries. Lancet Respir. Med. Comm. 2014;2:823–60.

91. WHO. WHO Guidelines for Indoor Air Quality: Household Fuel Combustion. Geneva, Switzerland; 2014.

92. Freeman OE, Zerriffi H. How You Count Carbon Matters: Implications of Differing Cookstove Carbon Credit Methodologies for Climate and Development Cobene fi ts. Environ. Sci. Technol. 2014;48:14112–20.

93. Ezzati M, Kammen DM, Ezzati M, Kammen DM. The Health Impacts of Exposure to and Data Needs The Health Impacts of Exposure to Indoor Air Pollution from Solid Fuels in Developing Countries : Knowledge , Gaps , and Data Needs. Environ. Health Perspect. 2002;110:1057–68.

94. Smith KR, McCracken JP, Weber MW, Hubbard A, Jenny A, Thompson LM, et al. Effect of reduction in household air pollution on childhood pneumonia in Guatemala (RESPIRE): a randomised controlled trial. Lancet [Internet]. Elsevier Ltd; 2011; 378:1717–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22078686

95. RTI International. Standard Operating Procedure for Particulate Matter (PM) Gravimetric Analysis. 2008.

96. Arku RE, Dionisio KL, Hughes AF, Vallarino J, Spengler JD, Castro MC, et al. Personal particulate matter exposures and locations of students in four neighborhoods in Accra, Ghana. J. Expo. Sci. Environ. Epidemiol. [Internet]. Nature Publishing Group; 2014;25:557–66.

97. Baumgartner J, Zhang Y, Schauer JJ, Ezzati M, Patz J a, Bautista LE. Household air pollution and children's blood pressure. Epidemiology [Internet]. 2012;23:641–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22659548

98. Dionisio K, Howie S, Dominici F, Fornace K, Spengler J, Adegbola R, et al. Household concentrations and exposures of children to particulate matter from biomass fuels in The Gambia. Environ. Sci. Technol. 2012;46:3519–27.

99. Lam N, Nicas M, Ruiz-Mercado I, Thompson LM, Romero C, Smith KR. Non-invasive measurement of carbon monoxide burden in Guatemalan children and adults following wood-fired temazcal (sauna-bath) use. J. Environ. Monit. [Internet]. 2011;13:2172–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21687856

100. Chisti MJ, Tebruegge M, La Vincente S, Graham SM, Duke T. Pneumonia in severely malnourished children in developing countries - mortality risk, aetiology and validity of WHO clinical signs: a systematic review. Trop. Med. Int. Health [Internet]. 2009;14:1173–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19772545

101. WHO, UNICEF. WHO child growth standards and the identification of severe

acute malnutrition in infants and children. 2009.

102. Barrera-Gomez J, Basagana X. Supplemental Material for " Models with transformed variables : interpretation and software " by J . Barrera-G o ´ mez and X . Basaga n. Epidemiology. 2015;26:e16–7.

103. Armstrong JR, Campbell H. Indoor air pollution exposure and lower respiratory infections in young Gambian children. Int. J. Epidemiol. 1991;20:424–9.

104. Clark ML, Peel JL, Burch JB, Nelson TL, Robinson MM, Conway S, et al. Impact of improved cookstoves on indoor air pollution and adverse health effects among Honduran women. Int. J. Environ. Health Res. [Internet]. 2009;19:357–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19626518

105. Fullerton DG, Bruce N, Gordon SB. Indoor air pollution from biomass fuel smoke is a major health concern in the developing world. Trans. R. Soc. Trop. Med. Hyg. 2008;102:843–51.

106. Ezzati M, Mbinda BM, Kammen DM. Comparison of Emissions and Residential Exposure from Traditional and Improved Cookstoves in Kenya. Environ. Sci. Technol. 2000;34:578–83.

107. Muller E, Diab RD, Binedell M, Hounsome R. Health risk assessment of kerosene usage in an informal settlement in Durban, South Africa. Atmos. Environ. 2003;37:2015–22.

108. WHO. WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide. 2006.

109. Walker CLF, Perin J, Katz J, Tielsch JM, Black RE. Diarrhea as a risk factor for acute lower respiratory tract infections among young children in low income settings. J. Glob. Health [Internet]. 2013; 3:010402. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3700029&tool=pmcentrez&rendertype=abstract

110. Fenn B, Morris SS, Black RE. Comorbidity in childhood in northern Ghana: magnitude, associated factors, and impact on mortality. Int. J. Epidemiol. [Internet]. 2005 [cited 2016 Jan 20];34:368–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15764695

111. Schmidt W-P, Cairncross S, Barreto ML, Clasen T, Genser B. Recent diarrhoeal illness and risk of lower respiratory infections in children under the age of 5 years. Int. J. Epidemiol. 2009;38:766–72.

112. Weisz A, Meuli G, Thakwalakwa C, Trehan I, Maleta K, Manary M. The duration of diarrhea and fever is associated with growth faltering in rural Malawian children aged 6-18 months. Nutr. J. [Internet]. 2011;10:25. Available from: http://www.nutritionj.com/content/10/1/25

113. O'Neill SM, Fitzgerald A, Briend A, Van den Broeck J. Child mortality as predicted by nutritional status and recent weight. J. Nutr. 2012;142:520–5.

114. Clasen T, Alexander K, Sinclair D, Boisson S, Peletz R, Chang H, et al. Interventions to improve water quality for preventing diarrhoea (Review). 2015

115. Schmidt W-P, Arnold BF, Boisson S, Genser B, Luby SP, Barreto ML, et al. Epidemiological methods in diarrhoea studies-An update. Int. J. Epidemiol. [Internet]. Schmidt, W.-P., Department for Disease Control, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom; 2011;40:1678–92. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L363091548

116. Fujii Y, Kaneko S, Nzou SM, Mwau M, Njenga SM, Tanigawa C, et al.

Serological surveillance development for tropical infectious diseases using simultaneous microsphere-based multiplex assays and finite mixture models. PLoS Negl. Trop. Dis. [Internet]. 2014 [cited 2014 Oct 22];8:e3040. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4117437&tool=pmcentrez&rendertype=abstract

117. Hanson KL, Cartwright CP. Use of an enzyme immunoassay does not eliminate the need to analyze multiple stool specimens for sensitive detection of Giardia lamblia. J. Clin. Microbiol. [Internet]. 2001;39:474–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=87761&tool=pmcentrez&rendertype=abstract

118. Moss DM, Priest JW, Boyd A, Weinkopff T, Kucerova Z, Beach MJ, et al. Multiplex bead assay for serum samples from children in Haiti enrolled in a drug study for the treatment of lymphatic filariasis. Am. J. Trop. Med. Hyg. [Internet]. 2011; 85:229–37. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3144818&tool=pmcentrez&rendertype=abstract

119. Priest JW, Bern C, Xiao L, Roberts JM, Kwon JP, Lescano AG, et al. Longitudinal analysis of Cryptosporidium species-specific immunoglobulin G antibody responses in peruvian children. Clin. Vaccine Immunol. [Internet]. Priest, J.W., Mail Stop F-13, Atlanta, GA 30341, United States; 2006;13:123–31. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L43437868

120. Priest JW, Li A, Khan M, Michael J, Lammie PJ, Ong CS, et al. Enzyme Immunoassay Detection of Antigen-Specific Immunoglobulin G Antibodies in Longitudinal Serum Samples from Patients with Cryptosporidiosis Enzyme Immunoassay Detection of Antigen-Specific Immunoglobulin G Antibodies in Longitudinal Serum Samples from. Clin. Diagn. Lab. Immunol. 2001;8:415–23.

121. Clasen T, Naranjo J, Frauchiger D, Gerba C. Laboratory assessment of a gravity-fed ultrafiltration water treatment device designed for household use in low-income settings. Am. J. Trop. Med. Hyg. 2009;80:819–23.

122. WHO. WHO guidelines on drawing blood : best practices in phlebotomy. World Heal. Organ. 2010;1–105.

123. WHO & UNICEF. WHO child growth standards and the identification of severe acute malnutrition in infants and children. A Jt. Statement by World Heal. Organ. United Nations Child. Fund. 2009.

124. Moss DM, Montgomery JM, Newland S V, Priest JW, Lammie PJ. Detection of cryptosporidium antibodies in sera and oral fluids using multiplex bead assay. J. Parasitol. 2004;90:397–404.

125. Corran PH, Cook J, Lynch C, Leendertse H, Manjurano A, Griffin J, et al. Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. Malar. J. [Internet]. 2008;7:195. Available from: http://www.malariajournal.com/content/7/1/195

126. Yuan Y. Sensitivity Analysis in Multiple Imputation for Missing Data. SAS Inst. Inc. 2014;1–12.

127. Spratt M, Carpenter J, Sterne JAC, Carlin JB, Heron J, Henderson J, et al. Strategies for multiple imputation in longitudinal studies. Am. J. Epidemiol. 2010;172:478–87.

128. Yuan YC. Multiple imputation for missing data: concepts and new developments. Inst. Tech. Rep. Rockville, MD; 2000.

129. Schafer J. Analysis of Incomplete Multivariate Data. New York, New York: Chapman & Hall/CRC; 1997.

130. Zeger S, Liang K-Y. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986;42:121–30.

131. Petersen MR, Deddens J a. A comparison of two methods for estimating prevalence ratios. BMC Med. Res. Methodol. [Internet]. 2008 [cited 2015 Jun 16];8:9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2292207&tool=pmcentrez&rendertype=abstract

132. Cox FE. Concomitant infections, parasites and immune responses. Parasitology [Internet]. 2001;122 Suppl:S23–38. Available from: http://www.journals.cambridge.org/abstract_S003118200001698X\nhttp://www.ncbi.nlm.nih.gov/pubmed/11442193

133. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): A prospective, case-control study. Lancet. 2013;382:209–22.

134. Barnard RJ, Jackson GJ. Giardia and Giardiasis: Biology, Pathogenesis, and Epidemiology. In: Erlandsen SL, Meyer EA, editors. Boston, MA: Springer US; 1984. p. 365–78. Available from: http://dx.doi.org/10.1007/978-1-4899-0594-9_21

135. Platts-Mills J a, McCormick BJJ, Kosek M, Pan WK, Checkley W, Houpt ER. Methods of Analysis of Enteropathogen Infection in the MAL-ED Cohort Study. Clin. Infect. Dis. [Internet]. 2014 [cited 2014 Nov 4];59 Suppl 4:S233–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25305292

136. Hanevik K, Hausken T, Morken MH, Strand EA, Morch K, Coll P, et al. Persisting symptoms and duodenal inflammation related to Giardia duodenalis infection. J. Infect. 2007;55:524–30.

137. Clasen T, Boisson S. Assessing the Health Impact of Water Quality Interventions in Low-Income Settings: Concerns Associated with Blinded Trials and the Need for Objective Outcomes. Environ. Health Perspect. 2015;

138. Barfod K. Estimation of Optimal Cutoff for ELISA Assays using Latent Class Methods and ROC Analysis. Proc. 11th Int. Symp. Vet. Epidemiol. Econ. [Internet]. 2006. Available from: www.sciquest.org.nz

139. Millar DG, Hirst TR, Snider DP. Escherichia coli Heat-Labile Enterotoxin B Subunit Is a More Potent Mucosal Adjuvant than Its Closely Related Homologue, the B Subunit of Cholera Toxin. Infect. Immun. 2001;69:3476–82.

140. Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, Van A, et al. The Global Enteric Multicenter Study (GEMS) of Diarrheal Disease in Infants and Young Children in Developing Countries : Epidemiologic and Clinical Methods of the Case / Control Study. 2012;55.

141. Williams RL. A Note on Robust Variance Estimation for Cluster-Correlated Data. Biometrics [Internet]. 2000;56:645–6. Available from: http://dx.doi.org/10.1111/j.0006-341X.2000.00645.x

142. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat. Methods Med. Res. 2007;16:219–42.

143. Fay R. Valid inferences from imputed survey data. Proc. Sect. Surv. Res. [Internet]. 1993;41–8.

144. Bouzid M, Hunter PR, Chalmers RM, Tyler KM. Cryptosporidium pathogenicity and virulence. Clin. Microbiol. Rev. 2013;26:115–34.

145. Faubert G. Immune response to Giardia duodenalis. Clin. Microbiol. Rev. 2000;13:35–54.

146. Heyworth MF. Immunology of Giardia and Cryptosporidium Infections. J. Infect. Dis. 1992;166:465–72.

147. Keselman A. Host and parasite factors contributing to variation in immunity and pathology in giardiasis. Georgetown University; 2014.

148. Maloney JG. Giardia lamblia, the Intestinal Microbiome, and Innate Immunity: A Study of the Host-Parasite Relationship during G. lamblia Infection. Georgetown University; 2015.

EMORY UNIVERSITY | Institutional Review Board

TO: Thomas Clasen, JD, PhD
Principal Investigator
Envir & Occup Health

DATE: April 21, 2014 (Correction to add Note)

RE: **Expedited Approval**
IRB00073615
Evaluation Phase 2: Assessing the health impact of advanced water filtration and improved cookstoves in Western Province, Rwanda: A cluster-randomized controlled trial

Thank you for submitting a new application for this protocol. This research is eligible for expedited review under 45 CFR.46.110 and/or 21 CFR 56.110 because it poses minimal risk and fits the regulatory categories F4, F5 and F7 as set forth in the Federal Register. The Emory IRB reviewed it by expedited process on 4/20/2014 and granted approval effective from **4/20/2014** through **4/19/2015**. Thereafter, continuation of human subjects research activities requires the submission of a renewal application, which must be reviewed and approved by the IRB prior to the expiration date noted above. Please note carefully the following items with respect to this approval:

- Study Protocol and Tools modified 3/18/2014
- Personal Contact Forms Phase 2 for Control Households, for Intervention Households and for Spillover Effects, all modified 3/16/2014
- Phase 2 Consent/ Information Sheets for Control Households, Intervention Households and Spillover Effects, all modified 3/16/2014
- Subpart D: Title 45 CFR 46.404/50.51. One parent's permission is required for participating minor children
- A waiver of consent is granted for Part 1 of the study only that uses administrative outcome data to assess overall effectiveness of the intervention.
- Information on previous study- Phase I design
- Information on previous study- RCT Phase I
- Information on previous study (sensors and compliance)
- NOTE: Emory IRB approval is contingent on full approval by the Rwanda National Ethics Committee.

Any reportable events (e.g., unanticipated problems involving risk to subjects or others, noncompliance, breaches of confidentiality, HIPAA violations, protocol deviations) must be reported to the IRB according to our Policies & Procedures at www.irb.emory.edu, immediately, promptly, or periodically. Be sure to check the reporting guidance and contact us if you have questions. Terms and conditions of sponsors, if any, also apply to reporting.

Before implementing any change to this protocol (including but not limited to sample size, informed

consent, study design), you must submit an amendment request and secure IRB approval.

In future correspondence about this matter, please refer to the IRB file ID, name of the Principal Investigator, and study title. Thank you

Regina Drake, M.Div, CIP
Senior Research Protocol Analyst
This letter has been digitally signed

CC: Zambrano Laura Envir & Occup Health

Emory University
1599 Clifton Road, 5th Floor - Atlanta, Georgia 30322
Tel: 404.712.0720 - Fax: 404.727.1358 - Email: irb@emory.edu - Web: http://www.irb.emory.edu/

EMORY UNIVERSITY | Institutional Review Board

TO: Thomas Clasen, JD, Ph.D.
Principal Investigator
Environmental & Occupational Health

DATE: April 20, 2015

RE: **Continuing Review Expedited Approval**
CR1_IRB00073615

IRB00073615
Evaluation Phase 2: Assessing the health impact of advanced water filtration and improved cookstoves in Western Province, Rwanda: A cluster-randomized controlled trial

Thank you for submitting a renewal application for this protocol. The Emory IRB reviewed it by the expedited process on 4/14/2015, per 45 CFR 46.110, the Federal Register expeditable categories (4), (5) and (7), and/or 21 CFR 56.110. This reapproval is effective from 4/14/2015 through **4/13/2016**. Thereafter, continuation of human subjects research activities requires the submission of another renewal application, which must be reviewed and approved by the IRB prior to the expiration date noted above. Please note carefully the following items with respect to this reapproval:

- Study Protocol and Tools, 5/28/2014
- Consent Form, Phase 2, 07/13/2014
- Phase 2 Control Consent Form, English + Kinyarwanda, 3/6/2014 uploaded 3/16/2014
- Phase 2 Spillover Consent Form, English + Kinyarwanda, 3/6/2014 uploaded 3/16/2014

Any reportable events (e.g., unanticipated problems involving risk to subjects or others, noncompliance, breaches of confidentiality, HIPAA violations, protocol deviations) must be reported to the IRB according to our Policies & Procedures at www.irb.emory.edu, immediately, promptly, or periodically. Be sure to check the reporting guidance and contact us if you have questions. Terms and conditions of sponsors, if any, also apply to reporting.

Before implementing any change to this protocol (including but not limited to sample size, informed consent, and study design), you must submit an amendment request and secure IRB approval.

In future correspondence about this matter, please refer to the IRB file ID, name of the Principal Investigator, and study title. Thank you.

Sincerely,

Regina Drake, M.Div, CIP
Senior Research Protocol Analyst
This letter has been digitally signed

CC: Zambrano Laura Envir & occup Health

Emory University
1599 Clifton Road, 5th Floor - Atlanta, Georgia 30322
Tel: 404.712.0720 - Fax: 404.727.1358 - Email: irb@emory.edu - Web: http://www.irb.emory.edu/
An equal opportunity, affirmative action university

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Dr. Thomas Clasen
DC / ITD
LSHTM

28 April 2014

Dear Dr. Clasen,

Submission Title: **ASSESSING THE HEALTH IMPACT OF ADVANCED WATER FILTERS AND IMPROVED COOKSTOVES IN WESTERN PROVINCE, RWANDA: A CLUSTER-RANDOMIZED CONTROLLED TRIAL**

LSHTM Ethics Ref: **7711**

Thank you for your response of 22 April 2014, responding to the Interventions Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval for the amendment having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Protocol / Proposal	Phase2 protocol (REVISED with tracked changes)_Combine.pdf	22/04/2014	22 April 2014
Information Sheet	INFORMATION SHEET FOR CONTROL PARTICIPANTS IN RWANDA RCT(T. Clasen 10 March 2014).docx	10/03/2014	10 March 2014

After ethical review

Any subsequent changes to the application must be submitted to the Committee via an Amendment form on the ethics online applications website. The Principal Investigator is reminded that all studies are also required to notify the ethics committee of any serious adverse events which occur during the project via an Adverse Event form on the ethics online applications website. An annual report form is required on the anniversary of the approval of the study and should be submitted during the lifetime of the study on the ethics online applications website. At the end of the study, please notify the committee via an End of Study form on the ethics online applications website. Ethics online applications website link: http://leo.lshtm.ac.uk

Yours sincerely,

Professor John DH Porter
Chair

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

Improving health worldwide

REPUBLIC OF RWANDA/REPUBLIQUE DU RWANDA

NATIONAL ETHICS COMMITTEE / COMITE NATIONAL D'ETHIQUE

Telephone: (250) 2 55 10 78 84
E-mail: info@rnecrwanda.org

Web site: www.rnecrwanda.org

Ministry of Health

P.O. Box. 84

Kigali, Rwanda.

FWA Assurance No. 00001973
IRB 00001497 of IORG0001100

May 02, 2014
No. 102/RNEC/2014

Dr. Thomas Clasen
Principal Investigator

Your Project title: **"ASSESSING THE HEALTH IMPACT OF ADVANCED WATER FILTERS AND IMPROVED COOKSTOVES IN WESTERN PROVINCE, RWANDA: A CLUSTER-RANDOMIZED CONTROLLED TRIAL"** has been evaluated by the Rwanda National Ethics committee.

Name	Institute	Involved in the decision		
		Yes	No (Reason)	
			Absent	Withdrawn from the proceeding
Dr.Jean-Baptiste MAZARATI	Biomedical Services (BIOS)	X		
Prof. Eugène RUTEMBESA	National University of Rwanda	X		
Dr.Laetitia NYIRAZINYOYE	National University of Rwanda(school of public Health)	X		
Prof.Alexandre LYAMBABAJE	National University of Rwanda	X		
Ms.Françoise UWINGABIYE	Lawyer at Musanze	X		
Dr. Egide KAYITARE	National University of Rwanda	X		
Sr.Domitilla MUKANTABANA	Kabgayi Nursing and Midwife school	X		

Mr. David K. TUMUSIIME	Kigali Health institute		X	
Dr. Lisine TUYISENGE	Kigali Teaching Hospital	X		
Dr. Claude MUVUNYI	Biomedical Services (BIOS)	X		

After reviewing your protocol **expedited review procedure** of 22 March 2014 **during** the where quorum was met, and revisions made on the advice of the RNEC submitted on 30 April 2014, **Approval has been granted to your study.**

Please note that approval of the protocol and consent form is valid for **12 months.**
You are responsible for fulfilling the following requirements:

1. Changes, amendments, and addenda to the protocol or consent form must be submitted to the committee for review and approval, prior to activation of the changes.
2. Only approved consent forms are to be used in the enrollment of participants
3. All consent forms signed by subjects should be retained on file. The RNEC may conduct audits of all study records, and consent documentation may be part of such audits.
4. A continuing review application must be submitted to the RNEC in a timely fashion and before expiry of this approval.
5. Failure to submit a continuing review application will result in termination of the study.
6. Notify the Rwanda National Ethics committee once the study is finished.

Sincerely,

RWANDA NATIONAL ETHICS COMMITTEE R.N.E.C P.O BOX:84 KIGALI

Dr. Jean- Baptiste MAZARATI
Chairperson, Rwanda National Ethics Committee.

Date of Approval: May 02, 2014
Expiration date: May 01, 2015

C.C.
- Hon. Minister of Health.
- The Permanent Secretary, Ministry of Health.

REPUBLIC OF RWANDA/REPUBLIQUE DU RWANDA

NATIONAL ETHICS COMMITTEE / COMITE NATIONAL D'ETHIQUE
Telephone: (250) 2 55 10 78 84
E-mail: info@rnecrwanda.org
Web site: www.rnecrwanda.org
Ministry of Health
P.O. Box. 84
Kigali, Rwanda.
FWA Assurance No. 00001973
IRB 00001497 of IORG0001100

August 11, 2014
No. 222/RNEC/2014

Thomas F. Clasen
Principal Investigator

Your Project title **"Amendment: Evaluation of DelAgua/ filter stove project in Rwanda-Phase2"** has been evaluated by the Rwanda National Ethics committee.

		Involved in the decision		
			No (Reason)	
Name	Institute	Yes	Absent	Withdrawn from the proceeding
Dr.Jean-Baptiste MAZARATI	Biomedical Services (BIOS)	X		
Prof. Eugène RUTEMBESA	University of Rwanda	X		
Dr.Laetitia NYIRAZINYOYE	University of Rwanda(school of public Health)	X		
Prof.Alexandre LYAMBABAJE	University of Rwanda	X		
Ms.Françoise UWINGABIYE	Lawyer at Musanze	X		
Dr. Egide KAYITARE	University of Rwanda	X		
Sr.Domitilla MUKANTABANA	Kabgayi Nursing and Midwife school	X		
Mr. David K. TUMUSIIME	Kigali Health institute	X		

Dr. Lisine TUYISENGE	Kigali Teaching Hospital		X	
Dr. Claude MUVUNYI	Biomedical Services (BIOS)	X		

After reviewing **amendments** to your protocol during the RNEC meeting of July 12, 2014 where quorum was met, and revisions made on the advice of the RNEC submitted on 05 August 2014, **Continuation of Approval has been granted to your study.**

You are responsible for fulfilling the following requirements:

1. Changes, amendments, and addenda to the protocol or consent form must be submitted to the committee for review and approval, prior to activation of the changes.
2. Only approved consent forms are to be used in the enrollment of participants
3. All consent forms signed by subjects should be retained on file. The RNEC may conduct audits of all study records, and consent documentation may be part of such audits.
4. A continuing review application must be submitted to the RNEC in a timely fashion and before expiry of this approval.
5. Failure to submit a continuing review application will result in termination of the study.
6. Notify the Rwanda National Ethics committee once the study is finished.

Sincerely,

RWANDA NATIONAL ETHICS COMMITTEE
R.N.E.C
P.O BOX:84 KIGALI

Dr. Jean- Baptiste MAZARATI
Chairperson, Rwanda National Ethics Committee.

C.C.
- Hon. Minister of Health.
- The Permanent Secretary, Ministry of Health

Republic of Rwanda

RWANDA BIOMEDICAL CENTER

A Healthy People. A Wealthy Nation

P.O. Box 84 KIGALI

National Health Research Committee
Ref: NHRC/2014/PROT/0163

To: **Dr. Fidel Ngabo**
Principal Investigator

Scientific Review Approval Notice

Dear Dr. Fidel Ngabo,

With reference to your request for approval of the Research Protocol entitled; **<< EVALUATION PHASE 2: ASSESSING THE HEALTH IMPACT OF ADVANCED WATER FILTERS AND IMPROVED COOKSTOVES IN WESTERN PROVINCE, RWANDA: A CLUSTER-RANDOMIZED CONTROLLED TRIAL>>**, We are pleased to inform you that, following a thorough review and critical analysis of your proposal (**Ref: NHRC/2014/PROT/0163 dated** 08/May/2014), your Research Protocol has been approved by National Health Research Committee.

However,

1) Changes amendments on approach and methodology must be submitted to the NHRC for review and approval to validate the changes.
2) A submission of quarterly progress report is mandatory
3) Submission to NHRC of final results before publication is mandatory
4) Failure to fulfill the above requirements will result in termination of study

Once again National Health Research Committee appreciates your interest in research and requests you to submit this proposal to the National Ethics Committee or IRB and then share a copy of the approval letter.

Your final approval reference number is **NHRC/2014/PROT/0163**

Yours Sincerely,

Dr. Parfait UWARIRAYE
Chairperson of NHRC

Signature:.....
Date:.....08/05/14

MATERIAL TRANSFER AGREEMENT

The PROVIDER (identified below) and the RECIPIENT (identified below) hereby agree to abide by all terms and conditions of the Uniform Biological Material Transfer Agreement ("UBMTA") published by the National Institutes of Health on March 8, 1995, attached as Appendix A hereto and incorporated by reference herein. To the extent supplies are available, the PROVIDER SCIENTIST (identified below) shall forward the material to the RECIPIENT SCIENTIST (identified below) upon full-execution of this Agreement. This Agreement is effective upon the date of last signature by the PROVIDER and RECIPIENT below.

Please fill in all of the blank lines below:

1. PROVIDER: Organization providing the ORIGINAL MATERIAL:

Organization:	Rwanda Biomedical Center/National Reference Laboratory
Address:	Kigali, Rwanda
MTA Contact Name:	Dr. Emil Ivan
MTA Contact Email Address:	emil.ivank@gmail.com

2. RECIPIENT: Organization receiving the ORIGINAL MATERIAL:

Organization:	Emory University
Address:	1599 Clifton Road NE, 4th Floor, Atlanta, GA USA 30322
MTA Contact Name:	Rajsekhar Guddneppanavar
MTA Contact Email Address:	mta@emory.edu

3. ORIGINAL MATERIAL (Enter description):
The material consists dried blood spots collected in connection with the biomarker and seroconversion sub-studies in connection with the field study led by the London School of Hygiene & Tropical Medicine to assess the impact of water filter and cook stoves in Western Province, Rwanda (RNEC approval no. 102, 14 May 2014). The samples consist of Whatman 903® dried blood spot "cards" and TropBio® dried blood spot filter paper "discs." Each card or disc is spotted with one to six drops of blood, approximately 20-80 uL each. All samples collected in connection with the study are included in this authorization. This is estimated to include up to 1600 samples.

4. Please describe the intended research to be conducted using the ORIGINAL MATERIAL:
The "cards" will be subject to multiplex analysis of various HAP exposure targets and biomarkers of acute lower respiratory infection (ALRI) and cardiovascular disease (CVD), as part of the "biomarker study." The "discs" will be subject to multiplex analysis of serological responses to various enteric pathogens and vaccines, as part of the "seroconversion study."

5. Termination date for this Implementing Letter (optional): _____Not applicable_____

6. Transmittal Fee to reimburse the PROVIDER for preparation and distribution costs (optional). Amount: _____

7. Selected Additional Terms attached as Appendix B hereto and incorporated by reference herein [ ] do / [ ] do not apply to this Agreement. If such Additional Terms apply, PROVIDER and RECIPIENT hereby agree to abide by such Additional Terms. In the event of conflict between such Additional Terms and other terms and conditions of this Agreement, such Additional Terms shall prevail.

Agreed and Accepted:

PROVIDER:

Authorized Official Signature: HEAD OF NATIONAL REFERENCE LABORATORY DIVISION CDC-KENYA EMORY REGIONAL CENTER

Signature Date:
Name: Emily Ivan
Title: Program Manager, NRL/PBR

GR

RECIPIENT:

Authorized Official Signature:

Signature Date: 9/9/15
Name: J. Cale Lennon, III
Title: Director, Licensing, Emory Office of Technology Transfer

Read and Acknowledged:

PROVIDER SCIENTIST

Signature:

Signature Date: 30/9/2015
Name: Dr Emily Ivan

By signing above, PROVIDER SCIENTIST understands that the material transferred pursuant to this agreement may be controlled under United States Export Controls regulations. PROVIDER SCIENTIST hereby certifies that to the best of PROVIDER SCIENTIST's knowledge, PROVIDER SCIENTIST will not export or re-export the material: i) to prohibited end-users or to embargoed destinations; ii) for use in any type of weapons proliferation activities, or iii) with knowledge that a violation of such regulations is about to occur (any questions regarding United States Export Controls regulations can be directed to the Office of Research Compliance at orc@emory.edu).

RECIPIENT SCIENTIST

Signature:

Signature Date: 8 September 2015
Name: Dr. Thomas F. Clasen
Email Address: thomas.f.clasen@emory.edu
Mailing Address for Materials: 1518 Clifton Road CNR 2nd Floor
Atlanta, GA USA 30322

0.1375B Centers for Disease Control and Prevention

Agreement to Prohibit CDC from Receiving Identifying Key

This agreement allows a non-CDC party that holds coded private information or human biological specimens to release the information or specimens to a receiving CDC investigator without CDC's becoming engaged in research involving human subjects. For basic instructions on using this form, see the next page.

1 Scope of agreement

A non-CDC party holds a key enabling linkage of identifying information to private information or specimens, and this key will not be released to the signing CDC investigator while the identifiable humans are alive.

The releasing non-CDC party attests that this release and the purposes of the planned research do not contradict the terms of consent under which the information or specimens were collected, whether that consent was documented or not documented. If the receiving CDC investigator learns the identity of one or more living individuals or, for previously unforeseen reasons, comes to believe that it is important to identify the individual(s), then this project might become research involving human subjects, and additional procedures must be followed to determine if the activity requires IRB approval under criteria at 45 CFR 46.111.

Brief description of information or specimens, such as protocol title(s) and reference number(s)

EVALUATION PHASE 2: ASSESSING THE HEALTH IMPACT OF ADVANCED WATER FILTERS AND IMPROVED COOKSTOVES IN WESTERN PROVINCE, RWANDA: A CLUSTER-RANDOMIZED CONTROLLED TRIAL / SEROCONVERSION SUB-STUDY

Additional comments

2 Signatures

Releasing non-CDC signatory

Signature — Date: 1 Oct 2015

Printed name: Thomas F. Clasen

(address): Rollins School of Public Health, 1518 Clifton Rd NE, Atlanta, GA

(voice) (fax)

(e-mail): TCLASEN@EMORY.EDU

Receiving CDC investigator

Signature — Date: Nov 3 2015

Printed name: Jeffrey W Priest

(address): 1600 Clifton Rd NE MS-D 66, Atlanta, GA 30329

(voice) (fax): (404) 718-4172

(e-mail): JPriest@cdc.gov

Secondary non-CDC signatory (optional)

Signature — Date

Printed name

Title

CDC science official

Signature — Date: 11/4/2015

Printed name: JEREMY SOBEL

Title: ASSOC. DIR. EPI SCIENCE, DFWED/CDC