Evaluating Objective Biological Measures to Characterize Enteropathogen Exposure

by

Leon Masunde Espira

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2021

Doctoral Committee:

Professor Joseph N.S Eisenberg, Chair
Associate Professor Veronica Berrocal, University of California, Irvine
Associate Professor Andrew Jones
Professor Nancy G. Love
Associate Professor Rafael Meza

Leon M. Espira

lespir@umich.edu

ORCID iD: 0000-0002-1320-3056

DEDICATION

To Darcy Hackert, for optimism to a future, and to my brother and sister, Anton Espira and Olga Espira for always being there for me.

ACKNOWLEDGEMENTS

I would like to begin by thanking my advisor, Dr Joe Eisenberg for talking with me all those years ago when I was first thinking about Michigan. Starting with a project on Q-fever in Kenya, my research has branched out more than I ever thought was possible. Most of all I would like to thank Joe for the independence and trust he put in me and for teaching me to think critically of everything that I put down on paper. Deciding on Ph.D advisor was one of the most pivotal decisions I have made in my career and I have could not have chosen better. I hope that Joe has enjoyed working me as much as I enjoyed working with him. I would also like to thank Dr. Nancy Love for opening her lab and department to me. None of this work would have been possible without her active support, engagement, and continued encouragement. I would like to thank Dr.Veronica Berrocal for keeping me grounded in statistical realities and not letting me be carried away in trying to use methods that my data would not support. I would also like to thank Dr. Andrew Jones for all our early conversations in getting my work in Ethiopia going. And thank you Dr. Rafael Meza for your kindness and support.

None of this research would have been possible without support of countless people in Ethiopia and Myanmar. In Ethiopia, I would like to thank Dr. Kaleab Baye for getting the study off the ground. I would also like to thank Brooke Gessesse, a fellow Ph.D student at Addis Ababa University- I hope we get to work together in the future. I would also like to thank Meseret Woldeyohannes from the Ethiopian Public Health Institute for helping me with lab space- he made a daunting task possible. I would like to thank Katrina Price and Taylor Lascko, two MPH students that I worked with in Ethiopia, for helping wrap up data collection. Finally, I

cannot thank my field team in Ethiopia enough for collecting the data and none of this would have been possible without the participants in Addis Ababa for selflessly taking part in my study.

I would like to thank Dr Pam Jagger and Dr Ther Aung for enabling me to go to Myanmar, a country that I hope to visit again one day. It was an eye-opening trip and my heart goes out to the people of Myanmar with the current situation as they strive to a better future.

I would like to thank my family for setting me on this path of discovery. It has not been the most straight forward of paths, or the easiest, but I forever remain indebted to my family for the sacrifices that have been made for me. To my brother Anton, for instilling a wonder about the world in me, to my sister Olga for being a rock for me, and to my parents, Susanna, and Abraham for all the opportunities they provided.

Lastly, I would like to thank all my funders. I am immensely grateful to the Dow Sustainability Program for funding me through two fellowships. I would also like to acknowledge Chinyere Neale from the Office of Global Public Health for all her help with funding. I would also like to thank Dr. Betsy Foxman and the Integrated Training in Microbial Systems Fellowship community for their help and support. Finally, I would like to acknowledge the Jacquez Koopman Simon Longini Complex Systems Epidemiology Scholarship for tuition support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The need for new metrics for the accurate assessment of Water, Sanitation, and Hygiene (WASH) interventions has been highlighted by the recent unexpected results from large-scale, cluster randomized intervention trials. Several reasons have been postulated about what may have caused these largely null results. In this dissertation, my primary focus is on the fact that these trials relied on exposure variables that measured the indirect impact of WASH interventions on pathogen loads and exposure and outcome variables based on self-report. A move towards exposure and outcome variables that are more specific may provide new insights into developing more targeted interventions. My secondary focus is on urban WASH, which has been understudied given that most WASH studies have been conducted in rural areas. Urban WASH is important because by 2050 68% of the world's population will reside in urban areas, often in informal settlements. These high-density informal settlements will present unique environmental risk factors. New metrics based on biological measures, such as pathogen quantification and assessments of physiological and immunological states will be necessary to inform the design of sustainable and effective interventions for urban areas. In this dissertation, I present three research aims focusing on improving our understanding of WASH in urban areas using biological measures.

Aim 1, examined the association between pathogen carriage in infants and biomarker derived scores. We used seven stool biomarkers to derive scores that measure gut integrity, acute inflammation, and chronic inflammation. The use of these biomarkers provides a novel way to gain insight on specific pathogenic processes when coupled with pathogen data. Score-pathogen

relationships were found to be enteropathogen-specific, reflecting expected physiological processes of epithelial invasion and gut dysfunction. Even though the biomarkers can be accurate measures of cellular damage and allow the evaluation of specific components of the inflammatory process, relationships between biomarkers are complex and depend on their cellular origin and the level of measurement. We postulate that there are multiple cellular damage processes that may differ between individuals and populations that have an end-state characterized by a run-away immune response, gut damage, and chronic malaise.

In Aim 2, we conducted a survey in urban informal settlements in Addis Ababa, Ethiopia to understand WASH indicators associated diarrheal prevalence in infants. We found that diarrheal prevalence was strongly associated with water treatment regime, and that boiling and water filtration were the most effective disease barriers. Our findings suggest that utility water in Addis Ababa is a major disease conduit and that systemic improvements are necessary to make utility water safe.

In Aim 3, we conducted a cross-sectional study, collecting and testing soil and water samples for a range of pathogens from urban and rural sites in and around Yangon, Myanmar. We quantified the concentrations of a range of viral, bacterial, and protozoan pathogens and one fecal indicator. We found that pathogen counts in both water and soil were significantly higher in urban areas compared to rural areas. We also found pathogen specific associations between environmental risks and environmental pathogen loads. Indicator and bacterial environmental loads were found to be associated with environmental risks, while environmental viral loads were not associated with any observed environmental risks. Our work highlights how environmental pathogen measurements provide added specificity towards identifying important

environmental pathways that require mitigation; and measuring pathogens provides additional specificity than measuring indicator organisms.

CHAPTER I

Introduction

1.1. Water, Sanitation and Health

I study diseases of poverty, their causes, and long-term sequelae. Diarrheal diseases and enteropathogen infections are the result of a lack of clean water and sanitation, the most basic resources that we all require to lead dignified and productive lives- clean water and sanitation. We have understood the utility and benefits of providing clean water and sanitation for more than 100 years. For example, after the introduction of water chlorination in England in 1911, infant mortality from diarrhea rapidly declined(1). However, despite the provision of clean water and sanitation being a central goal of both the Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs), billions still lack access to both. In 2017, it was estimated that 2.2 billion people lacked safely managed drinking water and 4.2 billion people lacked safely managed sanitation(2). Amid the COVID-19 pandemic, these gaps in access may have further implications, as personal hygiene, especially hand hygiene, is one of the most effective ways of preventing COVID-19(2).

The goal of water, sanitation, and hygiene (WASH) interventions is to stop susceptible individuals from coming into contact with enteropathogens, pathogens that are transmitted primarily via the fecal-oral route, and ultimately depleting or eliminating the environmental pathogen pool(3–6). The interactions between susceptible individuals and pathogens are illustrated by the well-known “F” diagram, which depicts both the fecal-oral transmission pathways and the barriers necessary to disrupt them(7). Sanitation interventions achieve this by

segregating human or animal excreta from environments that individuals come into the contact with. Water interventions aim to stop individuals from ingesting enteropathogens via their drinking water or domestic water used for food or bathing. Should any of these barriers fail, an individual may encounter a high dose of pathogens, and if ingested may result in infection which has acute (diarrheal disease) and chronic (growth faltering, cognitive impairment) outcomes.

1.1.1. Acute Outcomes of Pathogen Infection

Diarrhea is defined as the passage of 3 or more loose or watery stools in a 24-hour period and is a disruption of the absorptive capacity of the small intestine caused by the rapid transit of gastric contents(8,9). Diarrhea can be considered to be either osmotic or secretory(8). Osmotic diarrhea occurs when excessive numbers of osmotically active particles are present in the gut lumen, resulting in fluid passively moving into the lumen and exceeding the absorptive capacity of the gut, resulting in diarrhea(8). Secretory diarrhea occurs when the bowel mucosa secretes excessive amounts of fluid, either due to toxins such as the cholera toxin, or inherent abnormalities in enterocytes such congenital microvillous atrophy(8). Acute diarrhea is usually caused by infective agents that damage the mucosa, such as rotavirus or *Shigella*, or by pathogen toxins, as is the case with cholera(8). Chronic diarrhea typically lasts for longer than 3 weeks and is typically caused by protozoa parasites such as *Giardia* or *Cryptosporidium* or by innate defects to the mucosa such as damaged electrolyte transport systems or long-term damage of the enterocytes(8).

1.1.2. Chronic Outcomes of Pathogen Infection

Repeated enteric infections are also thought to result in abnormal small bowel physiology referred to as environmental enteric dysfunction (EED). EED has demonstrable histological features such as villous flattening, crypt hyperplasia, and lymphocytic infiltration of the lamina

propria coupled with chronic intestinal inflammation, that has been shown to have profound long-term effects, such as stunting and growth faltering(10–14). EED can compromise the integrity of the gut and allow potentially pathogenic bacteria to cross the gut wall, triggering a chronic inflammatory process(15). This inflammatory process may directly down-regulate growth, and indirectly affect growth through appetite suppression(16,17). Because EED compromises the integrity of the intestinal wall, an abnormally large number of white blood cells can infiltrate the gut, shunting an infant's metabolism to maintaining a chronically stressed immune surveillance system(18).

1.2. Measuring Exposure in WASH

Traditionally, the self-report of diarrhea over the past week or two has been used to measure the effectiveness of WASH interventions. The exposure measures that are measured using self-report are typically the ease of access to a defined sanitation technology. The idea is that if WASH interventions are effective in reducing host-pathogen contacts, either by reducing environmental pathogen loads or separating pathogens from hosts, fewer cases of diarrhea should occur during the defined time period. However, recent unexpected results from WASH trials have called into question the utility of standard metrics used to assess WASH effectiveness(19–25). There could be several reasons for these unexpected results; 1) these interventions did not cover enough of the community to significantly reduce pathogen transmission, and 2) the studies did not use appropriate metrics to measure the effects of the interventions. It may be more accurate to measure enteropathogen loads in both the environment and hosts or to evaluate biomarkers in subjects. The goal of any WASH intervention is better health outcomes, be it a reduction in diarrhea or improvements in long term outcomes, such as in child growth. Hence the need to accurately define metrics that are proximal to the exposure and provide an accurate

readout of infection or physiological outcomes. In the following two sections I will discuss the use of two alternative readouts for quantifying the effectiveness of WASH interventions: enteropathogen quantification and biomarker evaluation.

1.2.1. Enteropathogen Quantification

Enteropathogen quantification provides a direct measure of the dose of pathogens that individuals are exposed to. By providing a direct measure of exposure, enteropathogen quantification can provide direct readouts on how well a WASH intervention is working- the more effective an intervention the lower the concentration of pathogens that individuals are exposed to. Transmission pathway and pathogen specific measurements provide assessments on the relative importance of different transmission pathways and allow mitigation strategies to be tailored to achieve maximum effectiveness.

There are two approaches to evaluating enteropathogen exposure: external exposure assessments and internal exposure assessments(26). External exposure assessments quantify pathogens within an environmental substrate(26). External exposure assessments often rely on proxies, such as fecal indicators and source tracking markers, to measure enteropathogen loads along defined fecal-oral pathways(26). However, external assessments have two limitations: 1) they do not provide any measures on the dose of pathogen ingested, and 2) measurements derived from proxies may not accurately quantify the relative importance of the transmission pathways that they are supposed to measure(26). Depending on when they are obtained, internal exposure assessments can provide an estimate of ingested pathogen dose, typically via oral ingestion(26). However, these assessments provide no data on the actual transmission pathway and may be a measure of past exposure and not of current exposure(26).

The enumeration of enteropathogen loads can broadly be broken into three methods; culture-based methods, molecular methods, and metagenomics. Culture based methods typically begin with a filtration method to concentrate samples followed by plating on appropriate media. Culture based methods are limited by their low sensitivity and the presence of viable but not culturable pathogens(26). Molecular methods are typically based on the polymerase chain reaction (PCR). With new technology platforms such as Taqman Array Cards and digital PCR (dPCR) coming online, these techniques allow for the quantitative detection of multiple pathogens in a single sample(27–30). Molecular methods, however, remain resource intensive, requiring substantial laboratory infrastructure. Metagenomic techniques are based on the ability to sequence all the DNA present in a sample(31). The recent advent of Nanopore Sequencers has made this technique more accessible and adaptable to field conditions, allowing for the profiling of pathogen communities in environmental samples(32). Metagenomic techniques are however hampered by poor sensitivity in low abundance settings, the high cost of sequencing platforms, and a data analysis pathway that requires substantial bioinformatics expertise(26).

Despite the promise that direct enteropathogen quantification holds, as was pointed out by an Interdisciplinary Working Group, there are several inherent factors that make the accurate measurement of enteropathogen loads challenging(26). First, it is technologically difficult to quantify multiple pathogens whose pathogenicity and fate may differ by environmental conditions(26,33,34). Second, the measurement of environmental pathogen loads has to consider multiple fecal-oral transmission pathways, as commonly represented by the F-diagram(26,35). With multiple enteropathogen transmission pathways, it is important to accurately assess the relative importance of each transmission pathway and if pathways are pathogen specific(36).

Third, it is hard to accurately understand how environmental pathogen exposure affects host states such as gastrointestinal health, the microbiome or enteropathies such as EED(26,37,38).

The quantification of environmental enteropathogen loads was the focus of my third dissertation aim. I examined the association between environmental characteristics and pathogen loads in informal settlements and rural areas around Yangon, Myanmar.

1.2.2. Biomarker Measurements

The use of biomarkers to evaluate WASH effectiveness is relatively new, but several groups are working on developing biomarker and metabolic panels(39,40). To date, the main uses of biomarkers in the WASH field have been two-fold; 1) evaluating systemic inflammation responses to infection, and 2) measuring gut function to quantify the severity of EED(13). Widely used systemic inflammation biomarkers include c-reactive protein (CRP), α1acid glycoprotein and the insulin-like growth factor (IGF-1). There is a well-established body of work defining levels and thresholds for these biomarkers, making their use especially attractive. However, their utility is hampered by the fact that they represent broad reactions to physiological insults, making it hard to establish causal links between exposures and biomarkers. Additionally, with co-infections and asymptomatic pathogen carriage being the norm, attributing effects to specific pathogens is challenging, especially in small studies. Despite these limitations, biomarkers are useful for broad population surveillance.

A major potential use for using biomarkers as read outs for WASH interventions is that with a carefully selected panel of biomarkers, it should be possible to measure specific physiologic and metabolic pathways impacted by WASH interventions. The use of biomarkers may be more reflective of exposure than quantifying enteropathogens, given that some physiological responses are long-term and may better capture the long-term effectiveness of

WASH interventions. In addition, by examining the association between biomarkers and pathogen carriage, it may be possible to gauge pathogen specific physiological impacts, which may differ regionally. Making this link between pathogen carriage and biomarkers is the basis of my first dissertation aim. I focused on a panel of seven biomarkers and examined their association with stool pathogen loads. Three of the biomarkers that I used are widely deployed protein biomarkers used as a measure of EED; myeloperoxidase (MPO) alpha-1-antitrypsin (AAT) and neopterin(13,41,42). In addition, I used four recently described transcriptomic biomarkers; sucrase isomaltase (SI), caudal-homeobox-1 (CDX1), S100A8, and mucin-12(18,43–45). These seven biomarkers enable the measurement of specific cellular and immune processes such as cell damage and the neutrophil mediated innate immune response. The four transcriptomic biomarkers also enable the measurement of processes on shorter time scales and may be more indicative of recent or currently occurring physiological processes. In addition, given that the expression of the transcriptomic biomarkers is cell type specific, these biomarkers can be used to gauge the presence of different cell types in stool, reflecting cell type turnover in the intestine.

Myeloperoxidase: Myeloperoxidase is a cationic heme-containing enzyme found in primary azurophilic granules of neutrophils and, to a lesser degree, in primary lysosomes of monocytes(16). MPO is stored in cytoplasmic membrane-bound azurophilic granules and during stimulation, these granules are secreted out to the extracellular space by degranulation or exocytosis(17). Myeloperoxidase levels should therefore provide a readout of the neutrophil driven innate immune response, which is a rapid, non-specific response to pathogen infection.

Alpha-1-Antitrypsin: AAT is a water soluble, tissue diffusible, medium sized circulating glycoprotein with a blood half-life of 4-5 days(46). Over 80% of AAT is synthesized and

secreted by hepatocytes and the rest by monocytes, macrophages, the pancreas, lung alveolar cells, enterocytes, and the endothelium (46). The specific substrate of AAT is the serine proteinase elastase (46). In addition to inhibiting the excess of free elastase from neutrophils, AAT neutralizes a host of other enzymes including myeloperoxidase from neutrophils. Therefore, AAT should also act as a readout of the innate immune response, similarly to myeloperoxidase. In addition, AAT should enable us to gauge if the detected immune response is controlled or a detrimental runaway neutrophil response.

Neopterin: Neopterin is produced by activated monocytes, macrophages, dendritic cells, endothelial cells, and to a lesser extent in renal epithelial cells, fibroblasts, and vascular smooth muscle cells upon stimulation mainly by interferon gamma(IFNγ) (15). During acute viral infections, increased neopterin levels have been observed to correlate with disease severity(15). Neopterin is therefore a measure of the long-term adaptive immune response.

Sucrase Isomaltase: Sucrase Isomaltase is located on the brush border membrane (BBM) of the intestinal epithelium where it is involved in the final step of starch digestion(47). SI levels are reduced with mucosal injury but are relatively stable by race and age(48,49). SI activity is diminished in villus blunting (50), thus representing a possible surrogate marker of small intestinal function and integrity(51).

Caudal Homeobox-1: In humans there are three caudal homologues, Cdx1, Cdx2 and Cdx4(52). Due to the highly conserved homeodomain, Cdx1 and Cdx2 typically bind and transactivate many of the same DNA elements(52). SI was the first intestinal gene identified as a Cdx2 transcriptional target. Since Cdx1 is a close homologue of Cdx2, Cdx1 expression will track SI expression making Cdx1 another potential marker for gut function and integrity(52).

S100A8: S100A8 is part of the dimeric protein, S100A8/A9 (calprotectin). S100A8/A9 has been identified as a critical player of the inflammatory response. S100A8/A9 is found in the cytoplasm of neutrophils and monocytes and belongs to a large family of Ca^{2+} binding proteins(53). S100A8/A9 is part of the family of immune molecules known as endogenous ‘‘damage-associated molecular patterns’’ (DAMPs) that maintain a state of abnormal mucosal inflammation(54). DAMPs are liberated extracellularly, serving to signal danger to the host and promoting inflammation and repair processes that are initially beneficial and protective(54). However, in the setting of significant and persistent DAMP release, their downstream effects may result in tissue damage. DAMPs may have a central role in disease pathogenesis and provide another readout of the rapid innate immune response(54).

Mucin-12: The mucus layer overlying the epithelium promotes the elimination of gut contents and provides the first line of defense against physical and chemical injury caused by ingested food, pathogens, and microbial products. The major component of the mucus is secreted mucins, large glycoproteins with a highly polymeric protein backbone structure, linked to numerous hygroscopic and hydrophilic oligosaccharide side-chains that contribute to the formation of a gel-like structure(55). Mucin-12 is secreted by intestinal mucosal epithelial cells; therefore, the depletion of mucin-12 would be indicative epithelial cell damage(54).

1.3. WASH and Urbanization

My second and third dissertation aims focuses on understanding the larger picture of disease transmission in informal settlements. My work in Addis Ababa, Ethiopia examined how WASH impacted infant health in informal settlements. In Yangon Myanmar, I examined the association between the environmental site characteristics and pathogen loads to understand exposure associated environmental risks. Traditionally water and sanitation studies have been

carried out in rural areas. However, in a rapidly urbanizing world, understanding the impacts of water and sanitation interventions in urban areas is a priority(8). Though children in urban areas typically present with better long-term outcomes, they are more prone to short term acute insults, such as diarrhea, than rural infants. In addition, in urban areas, the issue may not be sanitation and water access as traditionally defined for mostly rural areas, but rather the quality of provided utilities and the ability of often underprepared public utilities to meet the demands of rapidly expanding populations in dense urban centers. This calls into question if interventions and metrics are transferable between urban and rural areas. Furthermore, urban and rural environments are likely to be dominated by different pathogens. The urban environment is likely to be dominated by more persistent pathogens such as viruses but, have pathogen hotspots that may drive risk. Rural areas may present a more stable environment, dominated by less persistent pathogens such as bacteria. Beside the differences in pathogens, modes of contact with pathogens may differ between urban and rural areas, with most pathogen contacts in urban areas occurring in the highly contaminated community environment while in rural areas most pathogen contact would occur in a contaminated home environment (29).

1.4. Dissertation Objectives

The overarching goal of my dissertation is to use biological measures to quantify pathogen exposure, both in the environment and from stool samples. Chapter 2 focuses on the physiological evaluation of pathogen carriage in infants from Addis Ababa, Ethiopia, by examining the association between stool pathogen loads and a panel of biomarkers. Chapter 3 focuses on understanding how water and sanitation exposures impact pathogen carriage and diarrheal prevalence in Addis Ababa, with a particular focus on understanding the role of utility provided water as potential driver of disease. In Chapter 4, I examine environmental

characteristics that drive pathogen contamination in informal settlements in Yangon and surrounding rural areas.

References

1. Woods RI, Watterson PA, Woodward JH. The causes of rapid infant mortality decline in England and Wales, 1861-1921. Part II. Popul Stud (NY). 1989;43(1):113–32.
2. The Sustainable Development Goals Report 2020. 2020.
3. Dangour AD, Watson L, Cumming O, Boisson S, Che Y, Velleman Y, et al. Interventions to improve water quality and supply, sanitation and hygiene practices, and their effects on the nutritional status of children. Cochrane Database Syst Rev. 2013;(8).
4. Strunz EC, Addiss DG, Stocks ME, Ogden S, Utzinger J, Freeman MC. Water, sanitation, hygiene, and soil-transmitted helminth infection: a systematic review and meta-analysis. PLoS Med. 2014;11(3).
5. Clasen TF, Alexander KT, Sinclair D, Boisson S, Peletz R, Chang HH, et al. Interventions to improve water quality for preventing diarrhoea. Cochrane Database Syst Rev. 2015;(10).
6. Freeman MC, Garn J V., Sclar GD, Boisson S, Medlicott K, Alexander KT, et al. The impact of sanitation on infectious disease and nutritional status: A systematic review and meta-analysis. Int J Hyg Environ Health. 2017;220(6):928–49.
7. Whyte LA, Jenkins HR. Pathophysiology of diarrhoea. Paediatr Child Heal. 2012;22(10):443–
8. Organization WH. Guidelines on sanitation and health. World Health Organization. 2018. 1–220 p.
9. Campbell RK, Schulze KJ, Shaikh S, Mehra S, Ali H, Wu L, et al. Biomarkers of environmental enteric dysfunction among children in rural Bangladesh. J Pediatr Gastroenterol Nutr. 2017;65(1):40–6.
10. Prendergast AJ, Humphrey JH. The stunting syndrome in developing countries. Paediatr Int Child Health. 2014;34(4):250–65.
11. Prendergast AJ, Rukobo S, Chasekwa B, Mutasa K, Ntozini R, Mbuya MNN, et al. Stunting is characterized by chronic inflammation in Zimbabwean infants. PLoS One. 2014;9(2).
12. Kosek M, Haque R, Lima A, Babji S, Shrestha S, Qureshi S, et al. fecal markers of intestinal inflammation and permeability associated with the subsequent acquisition of linear growth deficits in infants. Am J Trop Med Hyg. 2013;88(2):390–6.
13. Lee GO, Kosek P, Lima AAM, Singh R, Yori ÃPP, Olortegui MP, et al. Lactulose : Mannitol diagnostic test by hplc and of intestinal barrier function and environmental enteropathy. J Pediatr Gastroenterol Nutr. 2014;59(4):544–50.
14. Eisenhut M. Neopterin in diagnosis and monitoring of infectious diseases. J Biomarkers. 2013;2013.
15. Aratani Y. Myeloperoxidase : Its role for host defense , inflammation , and neutrophil function. Arch Biochem Biophys. 2018;640(November 2017):47–52.
16. Khan, Amjad A, Alsahli, Mohammed A, Rahmani AH. Myeloperoxidase as an active disease biomarker : recent biochemical and pathological perspectives. Med Sci. 2018;6(33).
17. Agapova S, Stephenson K, Manary M, Weisz A, Tarr PI, Mkakosya R, et al. Detection of low-concentration host mRNA transcripts in Malawian children at risk for environmental enteropathy. J Pediatr Gastroenterol Nutr. 2013;56(1):66–71.
18. Reese H, Routray P, Torondel B, Sinharoy SS, Mishra S, Freeman MC, et al. Assessing longer-term effectiveness of a combined household-level piped water and sanitation intervention on child diarrhoea, acute respiratory infection, soil-transmitted helminth infection and nutritional status: A matched cohort study in rural Odisha, . Int J Epidemiol. 2019;48(6):1757–67.

19. Pickering AJ, Djebbari H, Lopez C, Coulibaly M, Alzua ML, Bill F, et al. Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali : a cluster-randomised controlled trial. Lancet Glob Heal. 2015;3:e701–11.
20. Patil SR, Arnold BF, Salvatore AL, Briceno B, Ganguly S, Colford JM, et al. The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: A cluster randomized controlled trial. PLoS Med. 2015;11(8).
21. Clasen T, Boisson S, Routray P, Torondel B, Bell M, Cumming O, et al. Effectiveness of a rural sanitation programme on diarrhoea , soil-transmitted helminth infection , and child malnutrition in Odisha , India : a cluster-randomised trial. Lancet Glob Heal [Internet]. 2014;2(11):e645–53.
22. Humphrey JH, Mbuya MNN, Ntozini R, Moulton LH, Stoltzfus RJ, Tavengwa N V., et al. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. Lancet Glob Heal. 2019;7(1):e132–47.
23. Luby SP, Rahman M, Arnold BF, Unicomb L, Ashraf S, Winch PJ, et al. Effects of water quality , sanitation , handwashing , and nutritional interventions on diarrhoea and child growth in rural Bangladesh : a cluster randomised controlled trial. Lancet Glob Heal. 2018;6(3):E302–15.
24. Null C, Stewart CP, Pickering AJ, Dentz HN, Arnold BF, Arnold CD, et al. Effects of water quality , sanitation , handwashing , and nutritional interventions on diarrhoea and child growth in rural Kenya : a cluster-randomised controlled trial. Lancet Glob Heal. 2018;6(3):E316–29.
25. Goddard FGB, Ban R, Barr DB, Brown J, Cannon J, Colford JM, et al. Measuring Environmental Exposure to Enteric Pathogens in Low-Income Settings: Review and Recommendations of an Interdisciplinary Working Group. Environ Sci Technol. 2020;54(19):11673–91.
26. Lievens A, Jacchia S, Kagkli D, Savini C, Querci M. Measuring digital PCR quality: Performance parameters and their optimization. PLoS One. 2016;11(5):1–21.
27. Tsai K, Simiyu S, Mumma J, Aseyo RE, Cumming O, Dreibelbis R, et al. Enteric pathogen diversity in infant foods in low-income neighborhoods of Kisumu, Kenya. Int J Environ Res Public Health. 2019;16(3).
28. Liu J, Gratz J, Amour C, Nshama R, Walongo T, Maro A, et al. Optimization of quantitative PCR methods for enteropathogen detection. PLoS One. 2016;11(6):1–11.
29. Baker KK, Senesac R, Sewell D, Sen Gupta A, Cumming O, Mumma J. Fecal fingerprints of enteric pathogen contamination in public environments of Kisumu, Kenya, associated with human sanitation conditions and domestic animals. Environ Sci Technol. 2018;52(18):10263–74.
30. Tringe SG, Rubin EM. Metagenomics: DNA sequencing of environmental samples. Nat Rev Genet. 2005;6(11):805–14.
31. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol [Internet]. 2019;37(7):783–92.
32. Straub TM, Chandler DP. Towards a unified system for detecting waterborne pathogens. J Microbiol Methods. 2003;53(2):185–97.
33. Gupta S, Maiden MCJ. Exploring the evolution of diversity in pathogen populations. Trends Microbiol. 2001;9(4):181–5.
34. Wagner EG, Lanoix JN, Organization WH. Excreta disposal for rural areas and small communities. Vol. 39, World Health Organization. 1958.

35. Eisenberg JNS, Scott JC, Porco T. Integrating Disease Control Strategies : Balancing water sanitation and hygiene interventions to reduce diarrheal disease burden. Am J Public Health. 2007;97(5):846–52.
36. Harris VC, Haak BW, Boele van Hensbroek M, Wiersinga WJ. The intestinal microbiome in infectious diseases: the clinical relevance of a rapidly emerging field. Open Forum Infect Dis. 2017;4(3):1–8.
37. Kotton CN, Hohmann EL. Enteric pathogens as vaccine vectors for foreign antigen delivery. Infect Immun. 2004;72(10):5535–47.
38. Arndt MB, Cantera JL, Id LDM, Kalnoky M, White N, Bizilj G, et al. Validation of the Micronutrient and Environmental Enteric Dysfunction Assessment Tool and evaluation of biomarker risk factors for growth faltering and vaccine failure in young Malian children. PLoS Negl Trop Dis. 2020;1–24.
39. Giallourou N, Fardus-Reid F, Panic G, Veselkov K, McCormick BJJ, Olortegui MP, et al. Metabolic maturation in the first 2 years of life in resource-constrained settings and its association with postnatal growths. Sci Adv. 2020;6(15):1–11.
40. Colston JM, Yori PP, Colantuoni E, Moulton LH, Ambikapathi R, Lee G, et al. A methodologic framework for modeling and assessing biomarkers of environmental enteropathy as predictors of growth in infants: An example from a Peruvian birth cohort. Am J Clin Nutr. 2017;106(1):245–55.
41. McCormick BJJ, Lee GO, Seidman JC, Haque R, Mondal D, Quetz J, et al. Dynamics and trends in fecal biomarkers of gut function in children from 1-24 months in the MAL-ED study. Am J Trop Med Hyg. 2017;96(2):465–72.
42. Ordiz MI, Shaikh N, Trehan I, Maleta KM, Stauber J, Shulman R, et al. Environmental enteric dysfunction is associated with poor linear growth and can be identified by host fecal mRNAs. J Pediatr Gastroenterol Nutr. 2016;63(5):453–9.
43. Ordiz MI, Wold K, Kaimila Y, Divala O, Gilstrap M, Lu HZ, et al. Detection and interpretation of fecal host mRNA in rural Malawian infants aged 6–12 months at risk for environmental enteric dysfunction. Exp Biol Med. 2018;243(12):985–9.
44. Stauber J, Shaikh N, Ordiz MI, Tarr PI, Manary MJ. Droplet digital PCR quantifies host inflammatory transcripts in feces reliably and reproducibly. Cell Immunol. 2016;303:43–9.
45. Serres F De, Blanco I. Role of alpha-1 antitrypsin in human health and disease. 2014;311–35.
46. Gericke B, Amiri M, Naim HY. The multiple roles of sucrase-isomaltase in the intestinal physiology. Mol Cell Pediatr. 2016;3(1).
47. Ritchie BK, Brewster DR, Davidson GP, Tran CD, McNeil Y, Hawkes JS, et al. 13C-Sucrose Breath Test: novel use of a noninvasive biomarker of environmental gut health. Pediatrics. 2009;124(2):620–6.
48. Robayo-Torres CC, Diaz-Sotomayor ÃM, Hamaker BR, Baker SS, Chumpitazi BP, Opekun AR, et al. c-labeled-starch breath test in congenital sucrase-isomaltase deficiency. 2018;66(June):61–4.
49. Gupta S, Chong S, Fitzgerald J. disaccharidase activities in children : normal values and comparison based on symptoms and histologic changes. J Pediatr Gastroenterol Nutr. 1999;28(3):246–51.
50. Tooley KL, Saxon BR, Webster J, Zacharakis B, Mcneil Y, Davidson GP, et al. A novel non-invasive biomarker for assessment of small intestinal mucositis in children with cancer undergoing chemotherapy. Cancer Biol Ther. 2006;5(10):1275–81.

51. Guo RJ, Eun RS, Lynch JP. The role of Cdx proteins in intestinal development and cancer. Cancer Biol Ther. 2004;3(7):593–601.
52. Pruenster M, Vogl T, Roth J, Sperandio M. S100A8/A9: From basic science to clinical application. Pharmacol Ther. 2016;167:120–31.
53. Boyapati RK, Rossi AG, Satsangi J, Ho GT. Gut mucosal DAMPs in IBD: From mechanisms to therapeutic implications. Mucosal Immunol. 2016;9(3):567–82.
54. Kim YS, Ho SB. Intestinal goblet cells and mucins in health and disease: Recent insights and progress. Curr Gastroenterol Rep. 2010;12(5):319–30.

CHAPTER II[1]

Understanding the Physiological Impact of Pathogen Carriage in Infants using Biomarkers.

2.1. Introduction

Infant growth faltering remains a persistent concern in low- and middle-income countries (LMICs). Approximately 165 million children under 5 years of age were stunted based on the height-for-age Z-scores (HAZ) of – 2 or lower(1). It has long been recognized that enteric infections are an important cause of growth faltering; however, the mechanisms by which frequent pathogen infections and the physiological responses to these infections result in poorer growth remain poorly understood.

One potential mechanism that holds promise is that repeated enteric infections cause abnormal small bowel physiology resulting in a chronic condition referred to as environmental enteric dysfunction (EED). EED has demonstrable histological features such as villous flattening, crypt hyperplasia and lymphocytic infiltration of the lamina propria coupled with chronic intestinal inflammation(2–6), that have been shown to have profound long-term effects such as stunting and growth faltering. EED can compromise the integrity of the gut and allow potentially pathogenic bacteria to cross the gut wall, triggering a chronic inflammatory process(7). Inflammation may directly down-regulate growth, and indirectly affect growth

[1] Chapter II is in the process of revision for publication.

through appetite suppression(7,8). Because EED may compromise the integrity of the intestinal wall, an abnormally large number of white blood cells can infiltrate the gut, shunting an infant's metabolism to maintaining a chronically stressed immune surveillance system(7)(9,10).

The gold standard for diagnosis of EED is the histological examination of intestinal tissue. However, because of the invasive nature of collecting biopsies from very young children, there have been concerted efforts to develop alternative tests. The more widely accepted non-invasive tests are dual sugar absorption tests, with the most common one being the lactulose: mannitol (L:M) urine test. However, the L:M test is designed to measure intestinal permeability and provides no information on the immunological state of subjects(9,10). In addition, the L:M test is challenging to administer, and results are hard to compare across dosing and urine collection protocols and analytic platforms(11).

Other non-invasive biomarkers have been proposed. The most frequently reported panel is based on three protein stool biomarkers: alpha-1-antitrypsin (AAT), neopterin, and myeloperoxidase (MPO). Human alpha-1 antitrypsin is a water-soluble glycoprotein(12). During intestinal inflammation, AAT is extraverted into the gut and is therefore used as a biomarker of intestinal permeability and protein loss in EED(13). Neopterin is synthesized primarily by activated monocytes, macrophages, dendritic cells, and endothelial cells(14). It is therefore a marker of TH1 cell activation indicative of an adaptive immune response and has been used to diagnose autoimmune diseases such as celiac disease, considered a clinical and histopathological analog of EED(13). Finally, MPO is a marker of the neutrophil response. MPO has been correlated with disease activity in inflammatory bowel disease(13). These three protein stool biomarkers, however, do not provide the necessary specificity to understand the physiological

implications of EED, but rather provide a broad readout of immune responses that are often difficult to causally ascribe to a particular exposure.

In an effort to provide improved specificity, mRNA transcripts specific for intestinal inflammation have recently been investigated(15–18). Several features make these transcripts promising for the assessment of EED: 1) they are able to target and measure a range of intestinal processes related to the immune system much like AAT, neopterin, and MPO; 2) they are able to measure other non-immune processes such as epithelial state and nutrient absorption that are important indicators of EED; and 3) their use can be multiplexed, allowing for multiple markers to be measured simultaneously.

These mRNA transcripts have only been definitively described and measured for a single population of rural infants and children in Malawi, where they showed promise as indicators of EED based on their association with L:M test results and inflammatory indicators(15–18). There remain a number of questions about the use of these biomarkers for the assessment of EED. Of particular interest is whether these biomarkers accurately measure the features of EED that are most critical to child health, even if they are accurate measures of close histopathological analogs.

An additional challenge in diagnosing EED is that there is growing evidence that EED is not a single syndrome, but the result of multiple contributing processes that may differ geographically(19). Biomarkers should therefore be chosen to capture specific aspects of gut health, such as enterocyte damage, to adequately capture the most critical facets of EED. In addition, asymptomatic pathogen carriage is common among young children living without access to improved water and sanitation, and therefore, the use of biomarkers that measure immunological responses may not be the best strategy, since asymptomatic pathogen carriage

often results in transient immune activation(20,21). Asymptomatic, pathogen carriage still causes cellular and tissue damage, reinforcing the need for markers that provide cell and tissue specific measures of immune state and tissue integrity. It is also unclear what are the most relevant measures of child health, against which EED biomarkers should be evaluated. Currently, most studies rely on HAZ to evaluate EED, insofar as a biomarker that is associated with HAZ cross-sectionally, or predictive of changes in HAZ in the weeks-to-months following measurement, is considered 'valid'. However, growth faltering is a long term, multi-path process and establishing causal links between growth faltering and specific exposures is non-trivial. Biomarkers levels, especially at the transcriptome level, are often dynamic, necessitating robust study designs to adequately capture the physiological responses that drive biomarker levels. A recent study on the metabolome and its association with post-natal growth reported multiple metabolites associated with growth processes, but not metabolites associated with the physiological state of the gut(20). Similarly, a recently developed EED biomarker panel focused on evaluating micronutrient deficiencies and systemic infection but does not provide specific measures of gut integrity(22). Biomarkers that are associated with nutritional outcomes are not necessarily pertinent for the evaluation of EED or gut function, reinforcing the need for inclusion of gut specific biomarkers in panels, allowing for the evaluation of gut physiology at the tissue and cellular level.

In this chapter, we compare a panel of established fecal biomarkers (AAT, neopterin, and MPO) with 4 novel mRNA transcript biomarkers (sucrase isomaltase (SI), caudal homeobox 1 (Cdx1), S100A8 and mucin 12) to assess the added value of the novel biomarkers in informal settlements in Addis Ababa, Ethiopia. In this comparative analysis, two approaches were used. First, based on the literature we assign scores according to their biological role. Second, we used

data reduction methods to assign biomarkers to categories. The association between the derived scores and stool pathogen gene loads was then assessed using linear models to determine pathogen specific effects on gut physiology and attendant immune responses.

2.2. Materials and Methods

2.2.1. Study Population

Stool samples were collected in 2018 from 136 infants aged 6-23 months across 12 informal settlements in Addis Ababa, Ethiopia. Informal settlements for the survey were selected based on the following criteria: 1) they had not been designated for redevelopment and residents were not being relocated to housing developments at the time of the study; and 2) the selected settlements would only be redeveloped in 2025, when Ethiopia is scheduled to achieve middle-income status. Health Extension Workers were used to identify households with eligible infants. Households were then asked to respond to a survey on sanitation access, consent to the collection of anthropometric data and a stool sample.

2.2.2. Selection of 4 stool-based mRNA transcripts

We selected 4 stool mRNA transcripts to evaluate based on the ability: 1) to mirror the neutrophil response measured by MPO; and 2) to inform non-inflammatory processes, particularly the ability to evaluate gut integrity. The four mRNA transcripts selected were: sucrase isomaltase (SI), caudal homeobox 1 (Cdx1), S100A8, and mucin 12. SI is located on the brush border membrane (BBM) of enterocytes where it is involved in the final step of starch digestion and can therefore be considered a marker of gut integrity(23). SI levels are reduced with mucosal injury but are relatively stable by race and age(24). Caudal homeobox proteins are global transcription factors, and a close homologue of Cdx1 has been shown to transactivate SI,

and Cdx1 can therefore be used as a marker of gut integrity along with SI(25). S100A8 is part of the heterodimeric protein, S100A8/A9 (calprotectin) and is found in the cytoplasm of neutrophils and monocytes and is therefore a marker neutrophil response(26). Eosinophils also upregulate their expression of S100A8 and S100A9 in colonic inflammation and colonic repair(27). Mucin 12 is a membrane bound mucin constitutively expressed by enterocytes mainly in the colon and is therefore a potential marker of gut integrity along with SI and Cdx1(28,29).

The four transcriptomic biomarkers (SI, Cdx1, S100A8, and Mucin12) were selected for their ability to provide more specific cellular level readouts as transcriptome measurements are a measure of mRNA at the cellular level. Because these are cellular level measures, the presence of these biomarkers in stool is directly indicative of the presence of specific cell types in the stool. Though protein measurements are a better measure of gene expression and transcription, protein levels in stool are not indicative of the presence of cell types in stool. Specifically, SI, Cdx1, and Mucin12 mRNA transcripts are all indicative of the presence of enterocytes in the stool, with their expression levels being indicative of health of the enterocytes. S100A8 is indicates the presence of neutrophils in the stool and is therefore a measure of how many neutrophils have extravasated into the gut lumen, with higher S100A8 levels corresponding to the more acute infections. Given the measurement level, S100A8 may even be a more accurate measure of the neutrophil response than MPO.

2.2.3. Stool Sample Processing

After collection, stool samples were catalogued and scored for consistency. Samples were stored at 4°C overnight before processing the following morning. One aliquot of the stool sample, to be used for ELISA biomarker measurements, was stored at -80°C. A second aliquot

was processed for nucleic acid extraction using the ZymoBiomics DNA/RNA Mini Kit (Zymo Research, Irvine, CA).

2.2.4. ELISA biomarker measurements

ELISA's for three biomarkers (AAT, Neopterin, and MPO) were run on-site at the Ethiopian Institute of Public Health as per the manufacturer's instructions. AAT levels were determined using the Human Alpha-1-Antitrypsin ELISA from Biovendor Research and Diagnostic Products (Brno, Czech Republic) kit. Neopterin levels were determined using the GenWay Biotech Inc. (San Diego, USA) Neopterin ELISA. MPO levels were measured using the IDK® MPO ELISA from Immunodiagnostik AG (Bensheim, Germany).

2.2.5. Nucleic acid extraction

Standard kit protocol for nucleic acid isolation from fecal samples was used. Briefly, 200mg of stool was weighed out and placed in screw cap microcentrifuge tubes containing the DNA/RNA Shield Lysis Buffer. The tubes were packed in leak proof containers for transport to the University of Michigan where DNA and RNA were extracted as specified in the manufacturer's protocol within two months of sample collection. The quality and concentration of the extracted nucleic acids was measured using a Nanodrop Spectrophotometer (Thermo Scientific, Waltham, MA). The DNA and RNA were used to detect mRNA transcripts and pathogens.

2.2.6. Droplet Digital PCR Detection of mRNA Transcripts

Sample mRNA transcript loads were quantified using the QX200™ Droplet Digital™ PCR (ddPCR) system (Bio-Rad, Hercules, CA) using duplexed FAM and VIC TaqMan assays. Assays were setup using the One-Step RT-ddPCR Supermix (Bio-Rad, Hercules, CA) containing

5µl Supermix, 2µl Reverse Transcriptase, DTT at 15mM, 1µl each of each 20 xTaqMan Gene Expression Assay (Applied Biosystems, Carlsbad, CA) and 60ng of template or sterilized water to bring the reaction volume to 20µl. Droplets were generated using the QX200™ AutoDG™ Droplet Digital™ PCR system (Bio-Rad, Hercules, CA). 40µl of the generated droplets were loaded into a 96-well plate and sealed using a PX1™ plate sealer (Bio-Rad, Hercules, CA). Thermocycling conditions consisted of 60 min reverse transcription at 50°C, enzyme activation for 10 min at 95°C, followed by 40 cycles of denaturation at 94°C for 30 s, annealing and extension at 60°C for 1 min; followed by enzyme deactivation at 98°C for 10 min, and a continuous hold at 4°C. All samples were run on the C1000 Touch™ thermocycler (Bio-Rad, Hercules, CA). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) reactions were run separately for each sample for normalization of targets.

2.2.7. Droplet Digital PCR Detection of Stool Pathogens

Pathogen loads were quantified using the QX200™ Droplet Digital™ PCR system (Bio-Rad, Hercules, CA). DNA assays were setup by combining 10µl of ddPCR Supermix for Probes (no dUTP), primers at 900nM, probes at 250 nM and 60ng of template DNA or sterilized water for no template controls. RNA assays were setup using the One-Step RT-ddPCR Supermix (Bio-Rad, Hercules, CA) containing 5µl Supermix, 2µl Reverse Transcriptase, DTT at 15mM, primers at 900nM, probes at 250nM and 60ng of template RNA or sterilized water for no template controls. To set up rotavirus assays, samples were denatured at 95°C for 5 minutes and kept on ice prior to adding the RNA to the reaction. Droplets were generated using the QX200™ AutoDG™ Droplet Digital™ PCR system (Bio-Rad, Hercules, CA). 40µl of the generated droplets were loaded into a 96-well plate and sealed using a PX1™ plate sealer (Bio-Rad, Hercules, CA). Next the following 3 steps were taken: 1) For 40 cycles, DNA thermocycling

conditions consisted of enzyme activation at 95°C for 10 min; denaturation at 94°C for 30 s; annealing and extension at 58°C for 1 min. 2) Enzyme deactivation occurred at 98°C for 10 min; and 3) and a continuous hold at 4°C. RNA thermocycling conditions consisted of 1) 60 min reverse transcription at 50°C, 2) enzyme activation for 10 min at 95 °C, 3) 40 cycles of denaturation at 94°C for 30 s, 4) annealing and extension at 58°C for 1 min; 5) enzyme deactivation at 98°C for 10 min, and 6) a continuous hold at 4°C. All samples were run on the C1000 Touch™ thermocycler (Bio-Rad, Hercules, CA), with a lid temperature of 105°C, a sample volume of 40ul and a ramp rate of 2°C. On completion of the thermocycling, plates were read using the QX200™ Droplet Reader (Bio-Rad, Hercules, CA, USA) and QuantaSoft™ software (Bio-Rad, Hercules, CA). Samples were screened for seven bacteria (enteroaggregative *E.coli* (EAEC), enteropathogenic *E.coli* (EPEC), enterotoxigenic *E.coli* (ETEC), shigatoxigenic *E.coli* (STEC), *Shigella*, *Salmonella enterica*, and *Campylobacter spp*), three protozoans (*Entamoeba histolytica*, *Giardia lamblia*, and *Cryptosporidium spp*), three helminths (*Ancylstoma duodenale*, *Ascaris trichuris*, and *Stronyloides sterocalis*), and two viruses (rotavirus and norovirus GI and GII). EAEC, EPEC and ETEC were each screened for two gene targets.

2.2.8. Droplet Digital PCR Data Processing

All sample quantification was carried out using QuantaSoft™ software (Bio-Rad, Hercules, CA). Wells were checked and samples with <10,000 accepted droplets were rerun. To check the inter assay variability, 1/3 of the samples were randomly selected to be re-run for both mRNA transcript and pathogen quantification, and coefficients of variation (CV) were calculated. If more than 5% of re-run samples had CVs higher than 15%, all assays were rerun.

The quantification of mRNA transcripts was carried out by visually setting thresholds to distinguish between negative and positive droplets(15,16). The concentration in each well was then normalized to GAPDH and results presented as Target/GAPDH ratios.

To quantify stool pathogen loads, the threshold for differentiating negative from positive droplets was defined as one standard deviation above the negative droplet threshold on the no template controls(30,31). Following the setting of the threshold, all wells were visually inspected. Wells with less than three positive droplets were considered negative. Final concentrations of gene copies per 200mg of stool were obtained by multiplying by appropriate dilution factors. Final concentrations were then converted to log10. A value of 0.1 was added to zero values prior to log transformation.

2.2.9. Development of Inflammation Scores and Statistical Analysis

Three different immunological/physiological scores were developed for our panel of biomarkers: 1) a 'theory driven' score using our full biomarker panel based on a histological score framework similar to that described by Liu et al. (2020) (Table 2-1)(19); 2) a "data derived" principal component analysis (PCA) score based on the three commonly used EED biomarkers: AAT, MPO, and neopterin; and 3) a second "data derived" PCA score based on the full panel of biomarkers (Figure 2-1).

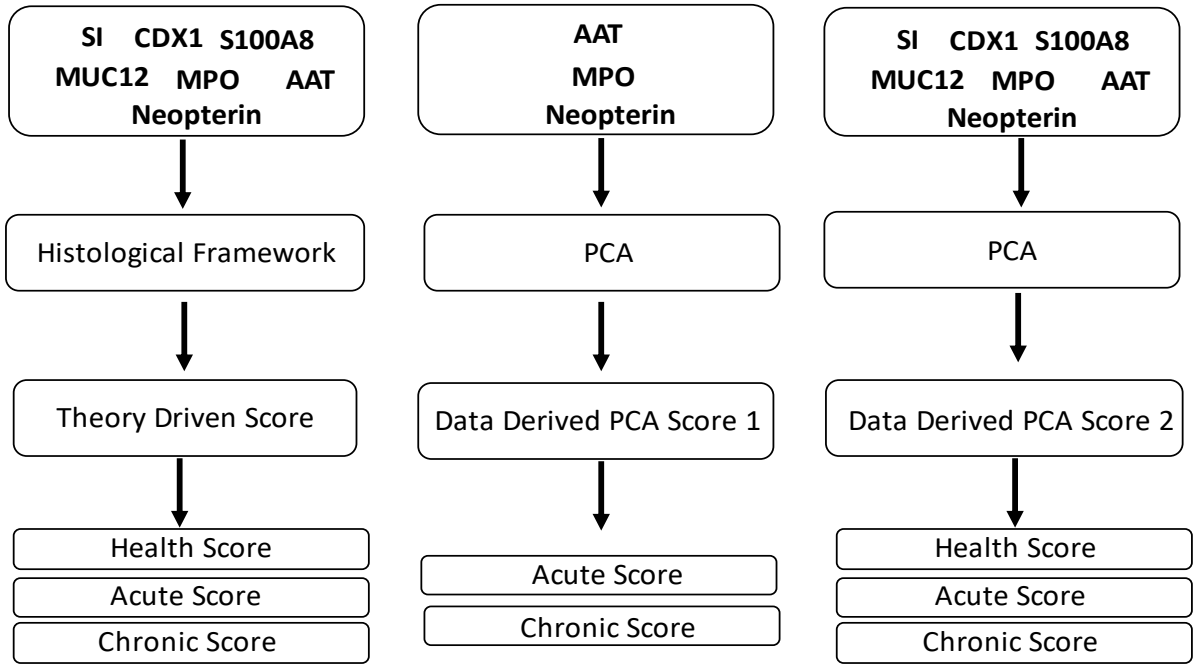


Figure 2-1: Framework for the derivation of the three biomarker scores presented in this chapter.

The 'theory driven' approach allows us to evaluate specific EED etiological processes based on the biological roles of selected biomarkers as reported in the literature. However, given the complexity of EED and poorly understood and complex relationships between the different biomarkers, this approach may subsume multiple etiologic processes into a single biomarker, especially with a small biomarker panel and sample. To obtain an infant's score, biomarkers were sorted into quintiles and each quintile was assigned a grade ranging from 0 to 4. Infants were then categorized into a quintile for each biomarker and assigned a grade corresponding their quintile. Three score types were derived for each infant within theory driven approach using different subsets of biomarkers: 1) a "Health Score" derived by summing SI, Cdx1 and mucin 12 grades for an infant. This score assessed enterocyte integrity. 2) An "Acute Inflammation Score" derived by summing an infant's S100A8, MPO and AAT grades. This score measured the neutrophil response; and 3) A "Chronic Inflammation Score" consisting of only of each infants neopterin grades, measuring the adaptive immune response (Figure 2-1).

The two “data driven” scores were both PCA derived (Figure 2-1). The ‘data driven’ approach allows the operationalization of relationships in the data that may be missed in the ‘theory driven’ approach. However, given the small sample size, spurious correlations in the data may influence the result. Given that a goal of our study was to evaluate whether it is possible to use fecal mRNA transcripts to measure gut specific processes, we did not specifically evaluate inter-biomarker associations. Future work with larger sample sizes will be needed to better evaluate any relationships between biomarkers. Prior to running PCAs, outliers were removed from the data and values standardized. Based on the Kaiser-Guttman rule and the Scree test, components were selected that explained at least 80% of the cumulative proportion of the variance. Factor loading values of > 0.4 or < -0.4 were used to select variables to be used in the creation of the score. An infant’s score was arrived at by summing the product of each biomarker value and its factor loading. Scores were then standardized around their means. The first data derived scores were only derived from MPO, AAT, and neopterin (Figure 2-1). The second series of data derived scores were derived from the full panel of biomarkers (Figure 2-1). Based on factor loading values, each series of scores consisted of an ‘Enterocyte Integrity’ score and inflammation scores that were classed as either, ‘Acute’, a measure the innate neutrophil response or ‘Chronic’, a measure of the adaptive immune response.

Correlation between biomarkers was assessed using Spearman correlation coefficients. The association between inflammation scores and stool pathogen loads was assessed using linear regression models. Models were always adjusted for the infant’s sex, age, and the consistency of the stool sample. All analysis was done using R version 4.0.3. PCA’s were performed using the FactoMineR package.

2.2.10. Ethical Approval

Ethical approval for sample collection and surveys was granted by institutional review boards at the University of Michigan (HUM00115103), and the Addis Ababa University (IRB/029/2017). Parents or legal guardians gave verbal, informed consent prior to participation or collection of data.

2.3. Results

2.3.1. Descriptive Statistics

Our sample of 136 were infants from 6 months to 2 years (mean = 14.35 months). Most were male (58.82%) and most were partially breastfed (77.21%). Only 6.61% of the infants were exclusively breastfeeding at the time of sampling. Most stool samples provided were formed (68.38%), with only 13.97% classified as liquid.

The mean LAZ and WAZ of the infants in our sample were -0.89 (SD=1.39) and -0.23 (SD=1.12) respectively, showing that stunting and wasting were not major concerns in our study population (Table 2-2).

MPO levels (mean=7080.21ng/mL, SD=12130.41) were measured in 134 stool samples, AAT levels (mean=596.42ng/mL, SD=617.67) were measured in 119 samples and neopterin levels (mean= 1509.34nmol/L, SD=1183.41) were measured in only 93 stool samples. Of the mRNA transcripts, S100A8 had the highest expression levels (mean= 10.68, SD=20.55) followed by mucin 12 (mean= 4.71, SD=6.29) and SI (mean= 2.62, SD=24.94). The transcription factor Cdx1 had the lowest expression levels (mean= 0.10, SD= 0.12) (Table 2-2). A comparison of our biomarkers values with previously reported studies is provided in Tables 2-3 to 2-9. In general, our values for MPO, AAT and neopterin were in-line with previously reported levels.

However, expression levels of 4 mRNA transcripts, SI, Cdx1, S100A8, and mucin 12 were notably higher than those reported in rural Malawian infants.

2.3.2. Spearman Correlation Coefficients between Biomarkers

The strongest significant correlation between biomarkers was between MPO and AAT (Spearman Coefficient= 0.45) (Table 2-10). MPO also had weak significant correlations with S100A8 (Spearman Coefficient = 0.23) and mucin 12 (Spearman Coefficient = -0.20). The transcription factor, Cdx1, was also weakly correlated with SI (Spearman Coefficient = 0.30) and S100A8 (Spearman Coefficient = -0.17). The correlation coefficients in our study were stronger than those previously reported.

2.3.3. Principal Component Analysis

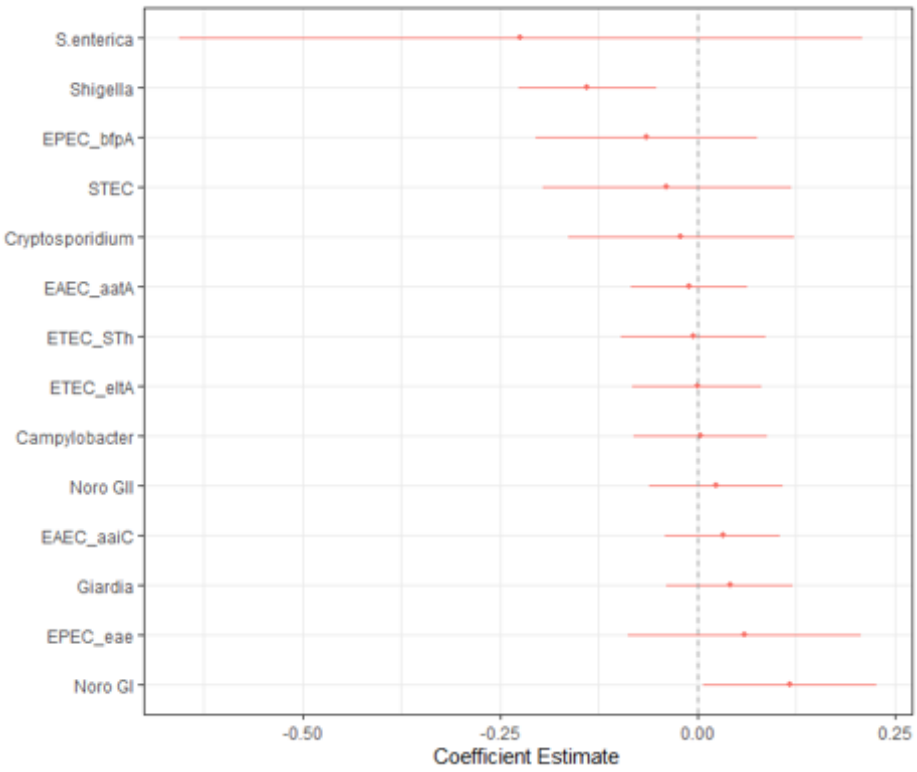
The first " data driven score" using the stool biomarkers AAT, MPO, and neopterin had two score types derived based on the factor loading values. The first is an **Acute Inflammation Score** measuring the neutrophil response with MPO (-0.69) and AAT (-0.64) having the highest factor loading values (Table 2-11). The second score is a **Chronic Inflammation Score** including only neopterin (-0.92).

The second "data driven score" utilizing the full panel of biomarkers allowed for five scores to be derived. The first were two chronic inflammation scores, **Chronic Inflammation A** score consisting of SI (0.44), AAT (0.49), and neopterin (0.50) and **Chronic Inflammation B** consisting of Cdx1(-0.41), mucin 12 (-0.63), and neopterin (0.62). Second, we had two acute scores, **Acute Inflammation A** consisting of Cdx1 (-0.40), S100A8 (0.46), and MPO (0.59) and **Acute Inflammation B** with Cdx1 (0.68) and S100A8 (0.49). The fifth was an **Enterocyte Integrity** score composed of SI (-0.59), mucin 12 (-0.47), and S100A8 (-0.49).

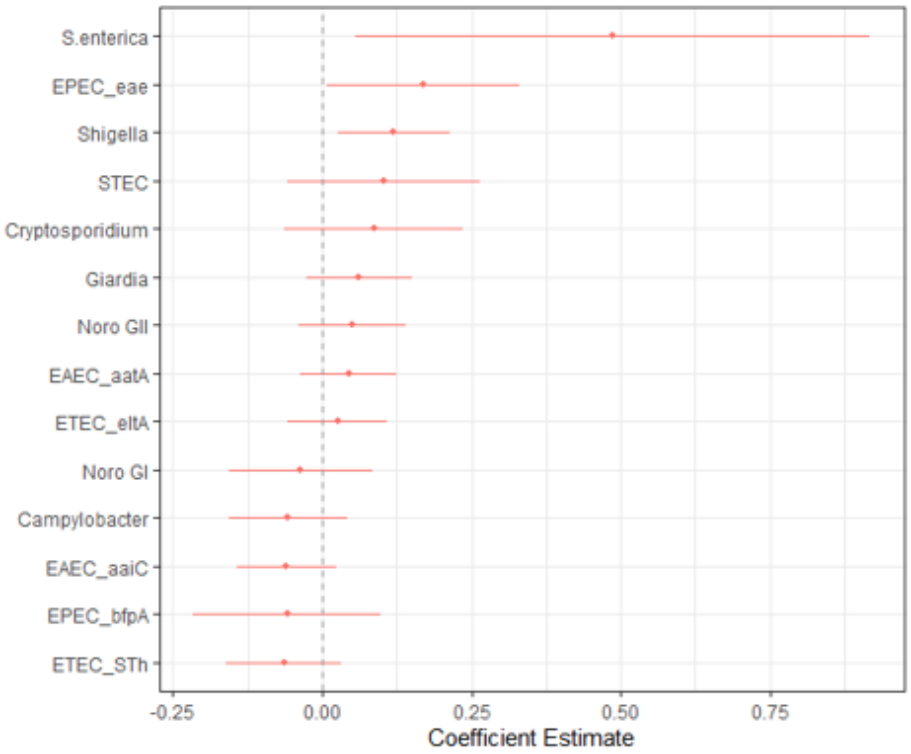
2.3.4. Associations between Stool Pathogen Loads and the Theory Derived Score

Linear regression models were used to examine association between stool pathogen loads and the 'theory' derived scores (Figure 2-2). In a linear regression model with the Enterocyte Integrity Score as the outcome and adjusted for all stool pathogens, infant age, breast feeding status and stool consistency, only *Shigella* was significantly associated with the score. Only *S.enterica* was significantly associated with the Acute Inflammation Score. The Chronic Inflammation Score was significantly associated with EAEC_*aaiC*, EAEC_*aatA*, and Noro_GI.

a. Pathogen Gene Counts and the Enterocyte Integrity Score



b. Pathogen Gene Counts and the Acute Inflammation Score



c. Pathogen Gene Counts and the Chronic Inflammation Score

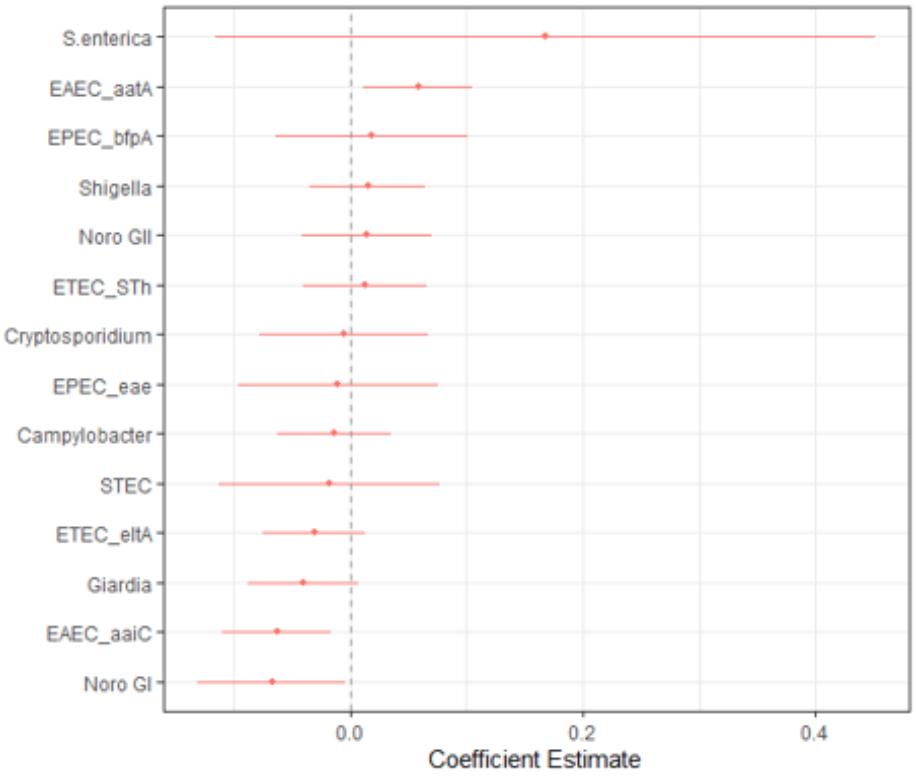
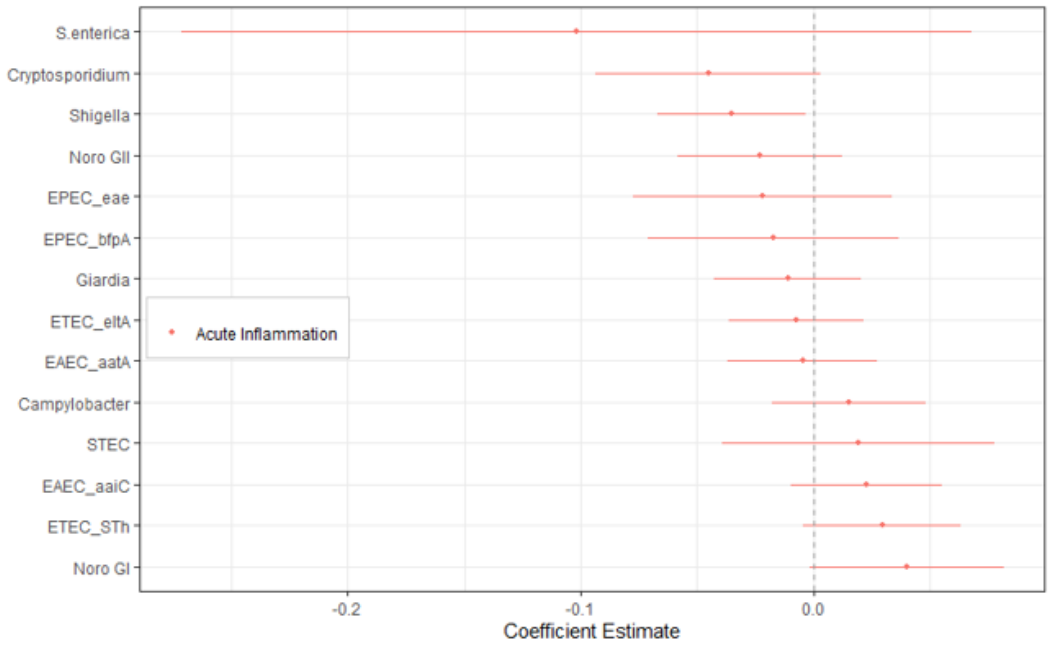


Figure 2-2: Associations between the theory derived score and stool pathogen gene counts; (a) associations between stool pathogen gene counts and the enterocyte integrity score, (b) associations between stool pathogen gene counts and the acute inflammation score, and (c) associations between stool pathogen counts and the chronic inflammation score.

2.3.5. Associations between Stool Pathogen Loads and the Data Derived PCA Score 1

Two scores were derived from the first data driven score consisting of MPO, AAT and neopterin (Figure 2-3). The Acute Inflammation score was only significantly associated stool *Shigella* gene counts. The Chronic Inflammation score was only associated with EAEC_*aaiC* and EAEC_*aatA*. Because of negative factor loading values, the associations were the opposite of what we noted in the other two scores.

a. Pathogen Gene Counts and the Acute Inflammation Score



b. Pathogen Gene Counts and the Chronic Inflammation Score

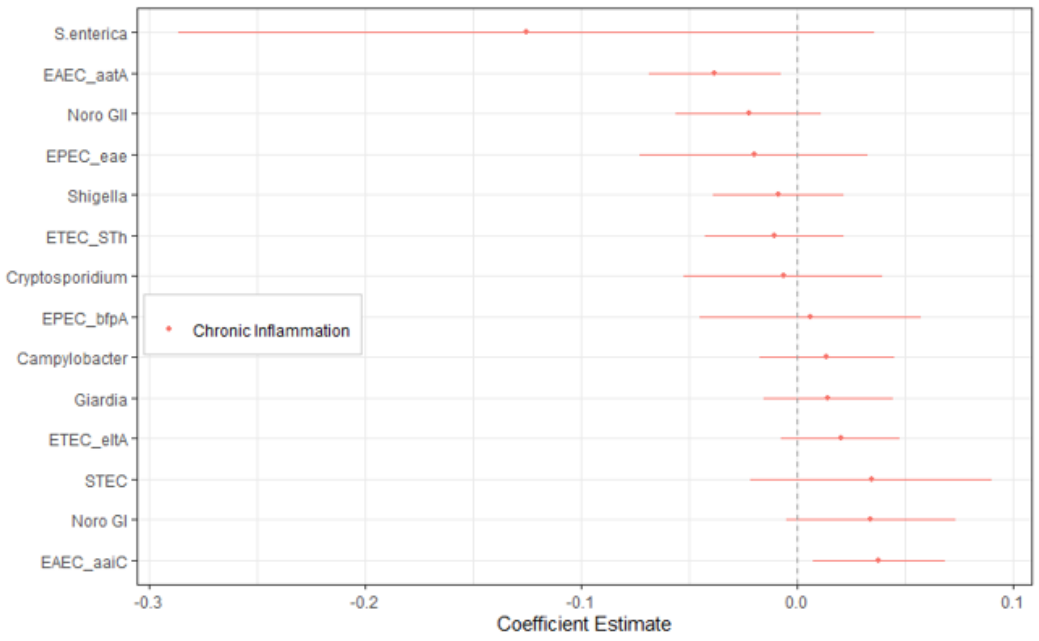
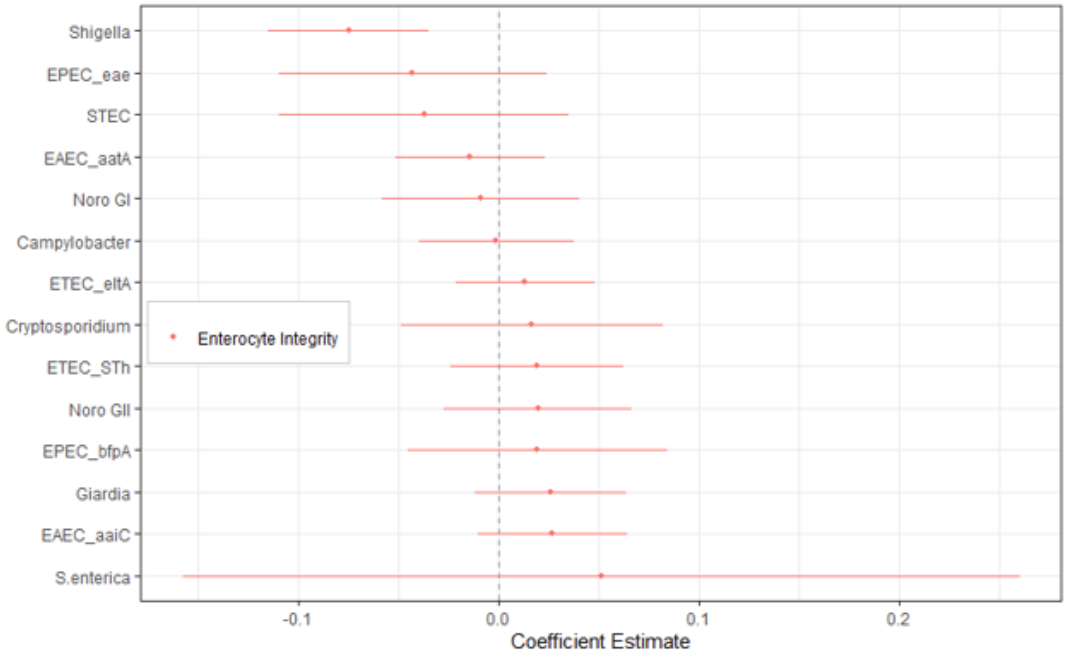


Figure 2-3: Associations between the ELISA derived scores and stool pathogen gene counts; (a) the association between pathogen gene counts and the acute inflammation score, and (b) the association between stool pathogen gene counts and the chronic inflammation score.

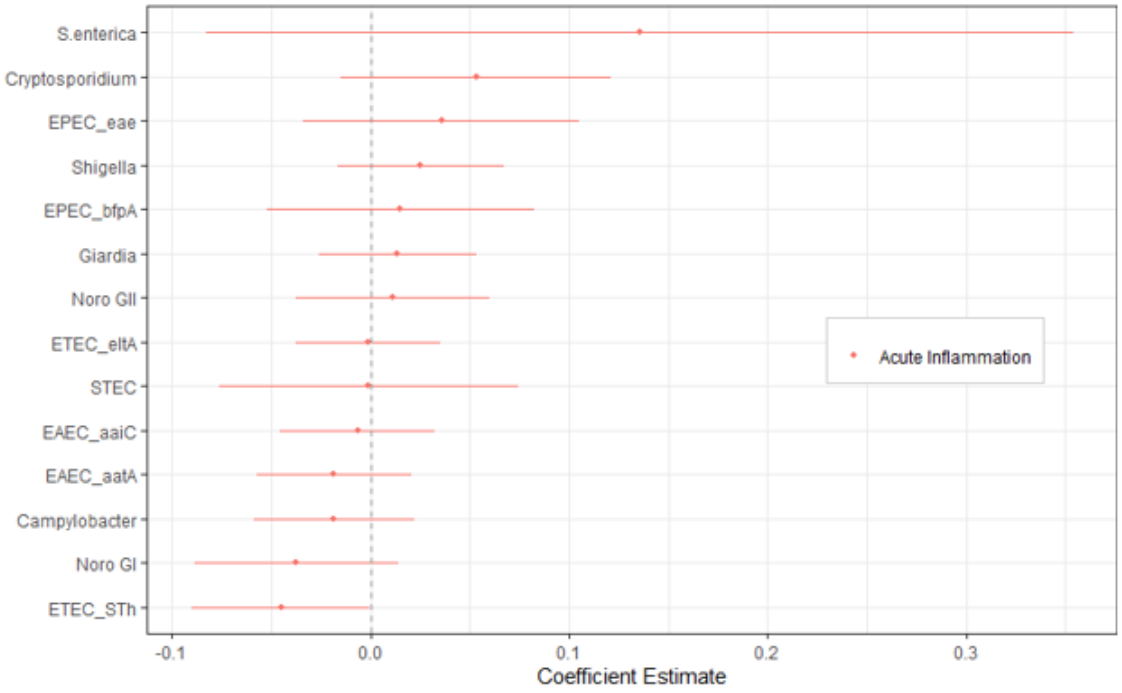
2.3.6. Associations between Stool Pathogen Loads and the Full Panel Score

Five scores were derived from the full panel data derived score 2 (Figure 2-4). Enterocyte Integrity was only significantly associated with *Shigella* stool gene counts. The Acute Inflammation A score did not have any significant associations at the 0.05 level, while the Acute Inflammation B score was only associated with *Shigella* gene counts. Chronic Inflammation A score was only associated with EAEC_*aaiC*, EAEC_*aatA*, and *S.enterica* stool gene counts. The Chronic Inflammation B score had significant associations with EAEC_*aaiC*, EAEC_*aatA*, and ETEC_*eltA*.

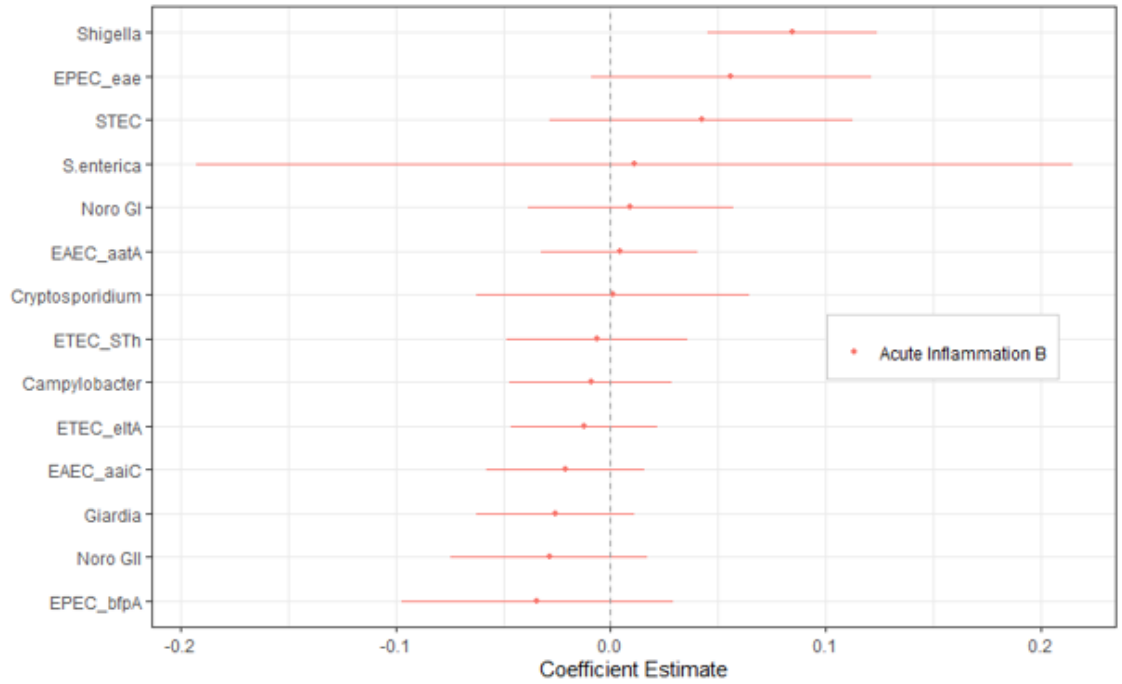
a. Pathogen Gene Counts and the Enterocyte Integrity Score



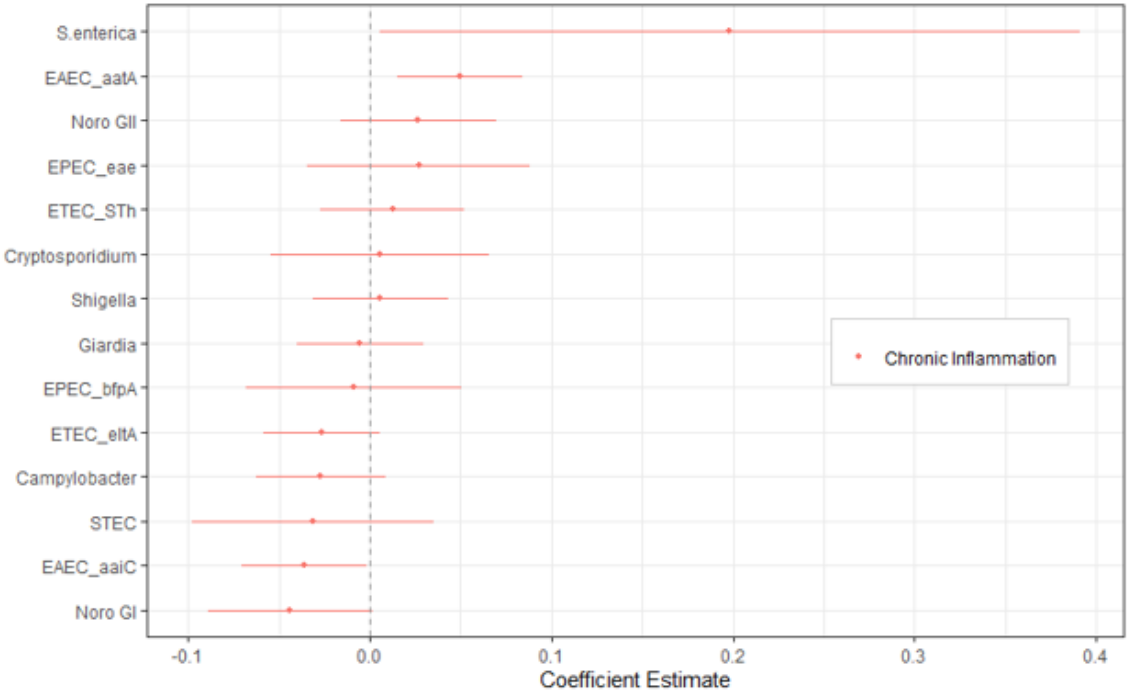
b. Pathogen Gene Counts and Acute Inflammation Score A



c. Pathogen Gene Counts and Acute Inflammation Score B



d. Pathogen Gene Counts and Chronic Inflammation Score A



e. Pathogen Gene Counts and Chronic Inflammation Score B

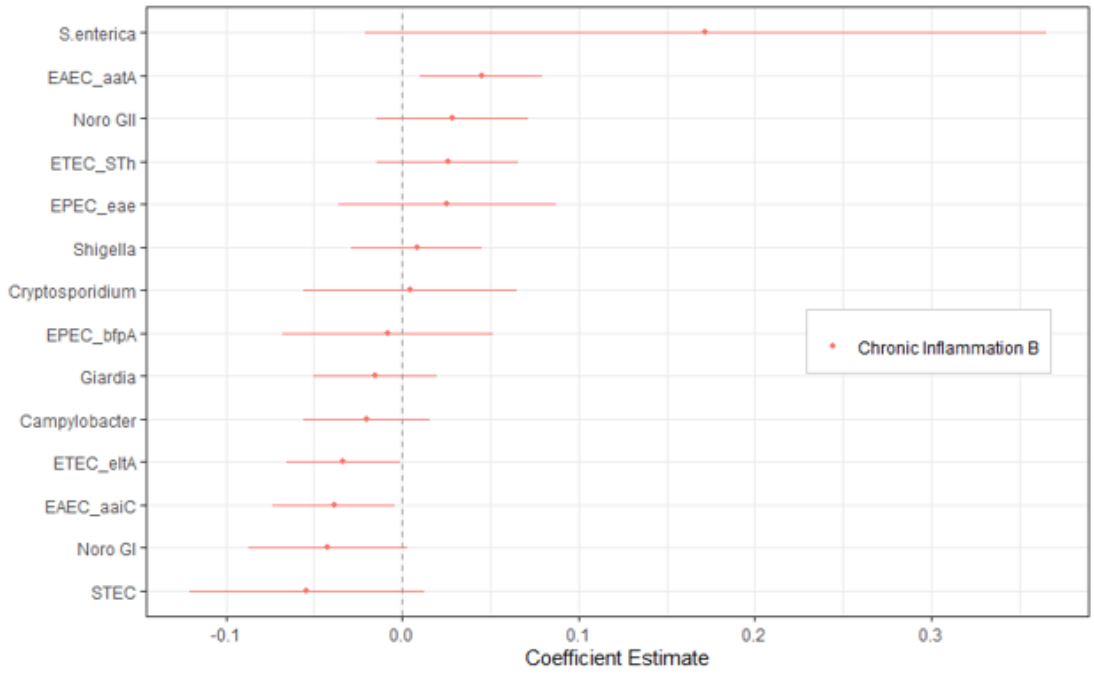


Figure 2-4: Associations between the full PCA data derived score and stool pathogen gene counts; a) association between the enterocyte integrity score and stool pathogen gene counts; b) association between the first acute inflammation score and stool pathogen gene counts; c) association between the second acute inflammation score and stool pathogen gene counts; d) association between the first chronic inflammation score and stool pathogen gene counts; and e) association between the second chronic inflammation score and stool pathogen gene counts.

2.4. Discussion

Our findings reinforce the concept that EED may be a non-specific syndrome that cannot be captured by a single catch-all definition but is rather the result of multiple overlapping responses to pathogen exposure, differing in time and physiological space(32). Nevertheless, the fact that multiple studies have linked existing biomarkers of EED, particularly MPO, to child growth, suggests that there is value in identifying biomarkers that can reliably characterize the facets of this syndrome that are most critical to child development(13,33). Our results also confirm the findings of Manary and colleagues in demonstrating that fecal mRNA transcripts have utility for the measurement of EED. They provide data in agreement with the data from fecal protein measurements, reflecting expected underlying biological processes. The utility of

fecal transcripts is the ability to measure specific cellular processes, providing a more complete picture of the overall gut immune process. Given the time scales that mRNA transcripts measure, they provide a more immediate snapshot of physiological and immune responses. They may be especially valuable in understanding dynamic processes such as pathogen infection and colonization. Given recent efforts to develop 'panels' of EED biomarkers(34,35), coupling panels with metabolic and transcriptomic approaches will be especially valuable.

Our 'theory' and 'data' derived scores largely agree. In both scores, biomarkers were grouped into three main score categories: enterocyte integrity, acute inflammation, and chronic inflammation. The inclusion of the fecal mRNA transcripts to score calculation provides more data on the systemic response to pathogen infection and colonization. Notably, the full panel score shows how *Shigella* infection not only elicits a strong neutrophil response, but also profoundly affects enterocyte integrity as shown by the negative association between *Shigella* gene stool loads and the Enterocyte Integrity score. We believe that the selected biomarkers detect the massive inflammatory response associated with apoptotic macrophages, infiltration of polymorphonuclear leukocytes and epithelial destruction that are characteristic of shigellosis(21). The inclusion of mRNA transcripts also informs on *S.enterica*, which utilizes macrophages to spread systemically. This co-opting of the macrophage response could explain the association of *S.enterica* with neopterin in addition to its impact on overall gut structure(36). Our scores may also illustrate how pathogenic *E. coli* carriage may be the norm in highly contaminated settings. Both EAEC and ETEC are associated with chronic inflammation processes, and the differences in the direction of the associations by pathotype markers may indicate how these *E.coli* pathotypes transition from acute effects during the initial infection process to a down regulation of the hosts immune system during biofilm formation and colonization(21,37).

Our mRNA transcript expression levels would indicate better gut health in our infants compared to infants in Malawi. The fecal protein biomarker measurements are in broad agreement with has previously been reported. The inflammation markers, MPO and neopterin, are marginally lower compared to other studies, while AAT levels were similar. AAT is a modulator of the neutrophil response and inhibits MPO, and the higher ratio of AAT to MPO in our sample compared to that reported in other studies, maybe indicative of a more controlled immune response in our samples(38). The differences in agreement between the mRNA transcript and stool protein markers may also indicate how mRNA transcripts and protein biomarkers may be measure responses that differ physiologically and temporally. Though at the protein level infants may be broadly similar, the differences between our study and previously reported studies at the transcript level may indicate fundamentally different responses by infants to different pathogenic stressors, but whose result is broad systemic inflammation. This reinforces the notion that EED, and systemic inflammation, are the endpoint of multiple physiological processes that may not only differ by region but also by the level of measurement. Specific biomarker panels to measure both specific cellular and physiologic processes and relevant EED endpoints need to be developed, rather than relying on panels that provide readouts of broad inflammatory end-states, whose causality is difficult to establish. The recently established causal relationship between duodenal microbiota and growth faltering will also help to identify metabolomic pathways to target for biomarker selection, both at the mRNA transcript and protein levels(6).

Overall, careful selection of biomarkers can help to inform our understanding of infection processes and how that these playout at the cellular and tissue levels. Biomarkers should be chosen to not only provide valid measures of long-term processes such as growth, but also

provide cell specific measures that help better understand the physiological responses to both transient insults and long-term asymptomatic pathogen carriage. Biomarker panels would ideally enable the differentiation of short-term physiological responses to transient infections from chronic end states such as EED. Alternatively, biomarker markers would enable the development of scales to identify where an infant is on what is likely a continuum of physiological responses, ranging from transient immune responses to chronic systemic dysregulation. The end goal of these biomarkers is to enable the precise evaluation of disease processes, provide information for therapeutic development and enable the development of lasting and sustainable solutions for the benefit of all.

Table 2-1: Derivation of the theory driven histological score based on indicators laid out in Liu et.al (2020).

Biomarker	Liu Indicators	Grading Scheme	Justification
Sucrase Isomaltase	Villus Architecture Gestalt Architecture Enterocyte Injury	Quintiles of SI levels with highest quintile corresponding to Grade 0 and lowest to Grade 4	Sucrase Isomaltase (SI) is located on the brush border membrane (BBM) of the intestinal epithelium where it is involved in the final step of starch digestion. SI is produced in the cytoplasm of epithelial cells before being transported and localized to the BBM surface. SI levels should be indicators of overall enterocyte structure and function
AAT	Epithelial Detachment Paneth Cell Density	Quintiles of AAT levels with the highest quintile corresponding to Grade 0 and the lowest to Grade 4	AAT is a measure of disturbed barrier function related to intestinal inflammation. The transmigration of neutrophils through the epithelial barrier promotes the release of host defense proteins eventually resulting in mucosal damage from chronic inflammation. The breakdown of epithelial barrier function results in the presence of serum proteins such as AAT in stool.
MUC12	Enterocyte Injury	Quintiles of Mucin 12 expression with the highest quintile corresponding to Grade 0 and the lowest to Grade 4	MUC12 is part of the transmembrane mucins that forms the glycocalyx and is produced by enterocytes. MUC12 expression decreases with enterocyte injury and should track with SI and CDX1.
CDX1	Enterocyte Injury	Quintiles of CDX1 expression with the highest quintile corresponding to Grade 0 and the lowest to Grade 4	CDX1 is closely tied to enterocyte structure and may be involved in the transcriptional regulation of SI
Neopterin	Intraepithelial Lymphocytes Chronic Inflammation of Lamina Propria	Quintiles of Neopterin expression with the highest quintile corresponding to Grade 0 and the lowest to Grade 4	Neopterin is a good marker of general inflammation as well as the presence of activated immune cells
MPO	Acute Inflammation	Quintiles of MPO expression with the highest quintile corresponding to Grade 0 and the lowest to Grade 4	MPO is marker of the short term rapid innate immune response which typically represents the first of line of defense against pathogens and injury. MPO is secreted by neutrophils
S100A8	Eosinophil Infiltration	Quintiles of S100A8 expression with the highest quintile corresponding to Grade 0 and the lowest to Grade 4	S100A8 is also a marker of acute inflammation and is secreted by both neutrophils and eosinophils depending on the healing/injury state.

Table 2-2: Selected demographics characteristics and descriptive statistics of infants and biomarker levels.

	N	Mean	SD	Minimum	Maximum	Median (25th, 75th percentiles)
LAZ	136	-0.89	1.39	-4.75	1.98	-0.89 (-1.90, 0.12)
WAZ	136	-0.23	1.12	-2.56	3.03	-0.38, (-1.03, 0.43)
MPO, ng/mL	134	7080.21	12130.41	137.40	102498.70	2970 (1382.9, 8563.4)
Neopterin, nmol/L	93	1509.34	1183.41	4.40	4635.60	1166.3 (576.1, 2176.4)
AAT, ng/mL	119	596.42	617.67	14.58	3915.67	411.64 (205.38, 753.33)
SI	136	2.62	24.94	0.00	288.18	0.027 (0.00, 0.087)
CDX1	136	0.10	0.12	0.00	0.69	0.070(0.027, 0.13)
MUC12	136	4.71	6.29	0.00	41.15	2.34 (1.15, 5.52)
S100A8	136	10.68	20.55	0.04	183. 17	4.48 (2.23, 13.45)

Table 2-3: Comparison of study Myeloperoxidase levels with previously reported levels.

Biomarker	Data Set	N	Mean +/- SD	Min	Max	Median (25th, 75th percentiles)
MPO(ng/mL)	**Ethiopian Infants**	**134**	**7080 +/- 12130.41**	**137.4**	**102498.7**	**2970 (1382.9, 8563.4)**
	Colston et al (2017)	4064	12482.8	0	111145.7	-
	Campbell et al (2017)	498	4460.3 (4145.0, 4799.5)*			
	Arndt et al (2016)	1185				3354.9 (1594.9, 7430.1)
	Kosek et al (2013)-Bangladesh	-	-	-	-	8838.23 (4552.95, 15020.98)
	Kosek et al (2013)-Brazil	-	-	-	-	6847.91 (3884.41, 12452.19)
	Kosek et al (2013)-India	-	-	-	-	14574.97 (6093.03, 27507.40)
	Kosek et al (2013)-Nepal	-	-	-	-	14484.40 (7499.47, 25317.29)
	Kosek et al (2013)-Peru	-	-	-	-	11623.52 (5765.75, 21883.94)
	Kosek et al (2013)-Pakistan	-	-	-	-	8452.01 (3950.82, 12868.78)
	Kosek et al (2013)-South Africa	-	-	-	-	16284.92 (6530.56, 25171.11)
	Kosek et al (2013)-Tanzania	-	-	-	-	17949.77 (9612.64, 2633.81)
	Kosek et al (2013)-Overall	-	-	-	-	11118.88 (5650.46, 20526.33)

* GM (95% CI)

Table 2-4: Comparison of study AAT levels with previously reported levels.

Biomarker	Data Set	N	Mean +/- SD	Min	Max	Median (25th, 75th percentiles)
AAT (ng/mL)	**Ethiopian Infants**	**119**	**596 +/- 617.67**	**14.58**	**3915.67**	**411.64 (205.38, 753.33)**
	Colston et al (2017)	4169	598.3	0.4	6337.8	-
	Kosek et al (2013)-Bangladesh	-	-	-	-	470 (250, 820)
	Kosek et al (2013)-Brazil	-	-	-	-	290 (140, 620)
	Kosek et al (2013)-India	-	-	-	-	590 (290, 1120)
	Kosek et al (2013)-Nepal	-	-	-	-	590 (310, 1120)
	Kosek et al (2013)-Peru	-	-	-	-	600 (310, 990)
	Kosek et al (2013)-Pakistan	-	-	-	-	230 (120, 0.52)
	Kosek et al (2013)-South Africa	-	-	-	-	250 (190, 630)
	Kosek et al (2013)-Tanzania	-	-	-	-	310 (180, 610)
	Kosek et al (2013)-Overall	-	-	-	-	440 (210, 860)

Table 2-5: Comparison of study Neopterin levels with previously reported levels.

Biomarker	Data Set	N	Mean +/- SD	Min	Max	Median (25th, 75th percentiles)
Neopterin (nmol/L)	**Ethiopian Infants**	**93**	**1509.34 +/- 1183.41**	**4.4**	**4635.6**	**1166.3 (576.1, 2176.4)**
	Colston et al (2017)	3892	2902.2	4.1	73505.9	-
	Campell et al (2017)	502	767.4 (716.5, 821.8)*			
	Arndt et al (2016)	1190				1017.6 (366.2, 2210.8)
	Kosek et al (2013)-Bangladesh	-	-	-	-	1422.35 (741.38, 2237.37)
	Kosek et al (2013)-Brazil	-	-	-	-	2385.39 (1895.55, 3303.96)
	Kosek et al (2013)-India	-	-	-	-	2009.31 (1417.08, 2969.13)
	Kosek et al (2013)-Nepal	-	-	-	-	1413.73 (1008.67, 2179.72)
	Kosek et al (2013)-Peru	-	-	-	-	1884.87 (1396.77, 2827.90)
	Kosek et al (2013)-Pakistan	-	-	-	-	2076.65 (1305.33, 3820.75)
	Kosek et al (2013)-South Africa	-	-	-	-	3997.17 (2492.34, 5830.18)
	Kosek et al (2013)-Tanzania	-	-	-	-	1748.40 (1155.42, 2960.52)
	Kosek et al (2013)-Overall	-	-	-	-	1846.68 (1171.39, 2997.85)

* GM (95% CI)

Table 2-6: Comparison of study transcript expression levels and expression levels previously reported in Malawian infants with varying L:M ratios.

Transcript	Ethiopian Infants		Agapova et al. (2013)			
			Normal L:M		Increased L:M	
	Mean	Median (25th, 75th percentiles)	Mean	Median	Mean	Median
SI	2.62	0.027 (0.00, 0.087)	0.0086	0.0103	0.008	0.5598
Cdx1	0.10	0.070(0.027, 0.13)				
S100A8	4.71	2.34 (1.15, 5.52)	0.7252	1.5854	0.2747	0.4888
Mucin 12	10.68	4.48 (2.23, 13.45)				

Table 2-7: Comparison of study transcript expression levels previously reported expression levels in Malawian infants.

Transcript	Ethiopian Infants		Stauber et al. (2016)	Ordiz et al. (2016)	
	Mean	Median (25th, 75th percentiles)	Median (25th, 75th percentiles)	Mean	Median (25th, 75th percentiles)
SI	2.62	0.027 (0.00, 0.087)	0.017 (0.008, 0.036)	0.114	0.017 (0.008, 0.036)
Cdx1	0.10	0.070(0.027, 0.13)	0.026 (0.016, 0.042)	0.047	0.027 (0.016, 0.042)
S100A8	4.71	2.34 (1.15, 5.52)	0.386 (0.154, 1.169)	0.979	0.386 (0.154, 1.169)
Mucin 12	10.68	4.48 (2.23, 13.45)	0.321 (0.162, 0.541)	0.447	0.294 (0.163, 0.539)

Table 2-8: Comparison of study transcript expression levels with expression levels in Malawian infants aged less than 12 months and 12-61 months.

Transcript	Ethiopian Infants		Ordiz et al. (2018)			
			Children <12 month		Children 12–61 months	
	Mean	Median (25th, 75th percentiles)	Mean	Median (25th, 75th percentiles)	Mean	Median (25th, 75th percentiles)
SI	2.62	0.027 (0.00, 0.087)				
Cdx1	0.10	0.070(0.027, 0.13)	0.024	0.018 (0.012, 0.029)	0.047	0.027 (0.016, 0.042)
S100A8	4.71	2.34 (1.15, 5.52)	1.927	1.406 (0.472, 2.304)	0.979	0.386 (0.154, 1.169)
Mucin 12	10.68	4.48 (2.23, 13.45)	0.351	0.217 (0.121, 0.450)	0.447	0.294 (0.163, 0.539)

Table 2-9: Comparison of study transcript expression levels with those of Malawian infants with severe EED and no to moderate EED.

Transcript	Ethiopian Infants		Ordiz 2018			
			Severe EED		Children no or moderate EED	
	Mean	Median (25th, 75th percentiles)	Mean	Median (25th, 75th percentiles)	Mean	Median (25th, 75th percentiles)
SI	2.62	0.027 (0.00, 0.087)				
Cdx1	0.10	0.070(0.027, 0.13)	0.02	0.018 (0.010, 0.025)	0.027	0.019 (0.013, 0.031)
S100A8	4.71	2.34 (1.15, 5.52)	2.09	1.33 (0.38, 2.36)	2.68	1.10 (0.50, 2.55)
Mucin 12	10.68	4.48 (2.23, 13.45)	0.306	0.192 (0.121, 0.393)	0.395	0.217 (0.123, 0.473)

Table 2-10: Spearman correlation coefficients between biomarkers. Coefficients significant at the 0.05 level are bolded.

	SI	S100A8	MUC12	CDX1	AAT	MPO	Neopterin
SI	1.00						
S100A8	-0.02	1.00					
MUC12	0.07	0.17	1.00				
CDX1	**0.30**	**-0.17**	0.08	1.00			
AAT	-0.12	0.02	0.04	-0.06	1.00		
MPO	-0.15	**0.23**	**-0.20**	-0.07	**0.45**	1.00	
Neopterin	-0.01	-0.19	-0.20	0.04	0.07	0.02	1.00

Table 2-11: Factor loading scores for the PCAs used to derive the two data driven scores.

	ELISA PCA		**Full Panel PCA**				
Biomarker	**PC1: Acute**	**PC2: Chronic**	**PC1: Chronic_A**	**PC2: Acute_A**	**PC3: Chronic_B**	**PC4: Healthy**	**PC5: Acute-B**
MPO	**-0.69**	0.08	0.32	**0.59**	-0.09	0.14	0.36
AAT	**-0.64**	0.38	**0.49**	0.33	-0.09	0.28	-0.27
Neopterin	-0.33	**-0.92**	**0.50**	0.00	**0.62**	0.01	0.07
SI			**0.44**	-0.39	0.15	**-0.59**	0.12
CDX1			0.22	**-0.40**	**-0.41**	0.31	**0.68**
MUC12			0.29	0.11	**-0.63**	**-0.47**	-0.28
S100A8			-0.28	**0.46**	0.10	**-0.49**	**0.49**
Standard deviation	1.24	0.98	1.22	1.16	1.08	0.99	0.92
Proportion of Variance	0.51	0.32	0.21	0.19	0.17	0.14	0.12
Cumulative Proportion	0.51	0.83	0.21	0.40	0.57	0.71	0.83

References

1. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, De Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet. 2013;382(9890):427–51.
2. Sullivan PB, Marsh MN, Mirakian R, Hill SM, Milla PJ, Neale G. Chronic diarrhea and malnutrition-histology of the small intestinal lesion. J Pediatr Gastroenterol Nutr. 1991;12:195–203.
3. Sullivan PB, Lunn PG, Clewes-Northrop C, Crowe PT, Marsh MN, Neale G. persistent diarrhea and malnutrition- the impact of treatment on small bowel structure and permeability. J Pediatr Gastroenterol Nutr. 1992;14:208–15.
4. Kelly P, Menzies IAN, Crane R, Zulu I, Nickols C, Feakins R, et al. Responses of small intestinal architecture and function over time to environmental factors in a tropical population. Am J Trop Med Hyg. 2004;70(4):412–9.
5. Campbell DI, Murch SH, Elia M, Sullivan PB, Sanyang MS, Jobarteh B, et al. Chronic T cell-mediated enteropathy in rural West African children: Relationship with nutritional status and small bowel function. Pediatr Res. 2003;54(3):306–11.
6. Chen RY, Kung VL, Das S, Hossain MS, Hibberd MC, Guruge J, et al. Duodenal microbiota in stunted undernourished children with enteropathy. N Engl J Med. 2020;383(4):321–33.
7. Crane RJ, Jones KDJ, Berkley JA. Environmental enteric dysfunction : An overview. Food Nutr Bull. 2015;36(1):76–87.
8. Sanderson IR. Growth problems in children with IBD. Nat Publ Gr. 2014;11(10):601–10.
9. Campbell RK, Schulze KJ, Shaikh S, Mehra S, Ali H, Wu L, et al. Biomarkers of environmental enteric dysfunction among children in rural Bangladesh. J Pediatr Gastroenterol Nutr. 2017;65(1):40–6.
10. Prendergast AJ, Humphrey JH, Mutasa K, Majo FD, Rukobo S, Govha M, et al. Assessment of environmental enteric dysfunction in the SHINE Trial: Methods and challenges. Clin Infect Dis. 2015;61(November):S726–32.
11. Lee GO, Kosek P, Lima AAM, Singh R, Yori ÃPP, Olortegui MP, et al. Lactulose : Mannitol diagnostic test by hplc and of intestinal barrier function and environmental enteropathy. J Pediatr Gastroenterol Nutr. 2014;59(4):544–50.
12. Serres F De, Blanco I. Role of alpha-1 antitrypsin in human health and disease. 2014;311–35.
13. Kosek M, Haque R, Lima A, Babji S, Shrestha S, Qureshi S, et al. Fecal markers of intestinal inflammation and permeability associated with the subsequent acquisition of linear growth deficits in infants. Am J Trop Med Hyg. 2013;88(2):390–6.
14. Eisenhut M. Neopterin in diagnosis and monitoring of infectious diseases. J Biomarkers. 2013;2013.
15. Agapova S, Stephenson K, Manary M, Weisz A, Tarr PI, Mkakosya R, et al. Detection of low-concentration host mRNA transcripts in Malawian children at risk for environmental enteropathy. J Pediatr Gastroenterol Nutr. 2013;56(1):66–71.
16. Stauber J, Shaikh N, Ordiz MI, Tarr PI, Manary MJ. Droplet digital PCR quantifies host inflammatory transcripts in feces reliably and reproducibly. Cell Immunol. 2016;303:43–9.
17. Ordiz MI, Shaikh N, Trehan I, Maleta KM, Stauber J, Shulman R, et al. Environmental enteric dysfunction is associated with poor linear growth and can be identified by host fecal mrnas. J Pediatr Gastroenterol Nutr. 2016;63(5):453–9.

18. Ordiz MI, Wold K, Kaimila Y, Divala O, Gilstrap M, Lu HZ, et al. Detection and interpretation of fecal host mRNA in rural Malawian infants aged 6–12 months at risk for environmental enteric dysfunction. Exp Biol Med. 2018;243(12):985–9.
19. Liu TC, Vanbuskirk K, Ali SA, Kelly MP, Holtz LR, Yilmaz OH, et al. A novel histological index for evaluation of environmental enteric dysfunction identifies geographic-specific features of enteropathy among children with suboptimal growth. PLoS Negl Trop Dis. 2020;14(1):1–21.
20. Yu J, Ordiz MI, Stauber J, Shaikh N, Trehan I, Barnell E, et al. Environmental enteric dysfunction includes a broad spectrum of inflammatory responses and epithelial repair processes. Cmgh. 2016;2(2):158-174.e1.
21. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent advances in understanding enteric pathogenic Escherichia coli. Clin Microbiol Rev. 2013;26(4):822–80.
22. Arndt MB, Cantera JL, Id LDM, Kalnoky M, White N, Bizilj G, et al. Validation of the Micronutrient and Environmental Enteric Dysfunction Assessment Tool and evaluation of biomarker risk factors for growth faltering and vaccine failure in young Malian children. PLoS Negl Trop Dis. 2020;1–24.
23. Gericke B, Amiri M, Naim HY. The multiple roles of sucrase-isomaltase in the intestinal physiology. Mol Cell Pediatr. 2016;3(1).
24. Robayo-Torres CC, Quezada-Calvillo R, Nichols BL. Disaccharide digestion: Clinical and molecular aspects. Clin Gastroenterol Hepatol. 2006;4(3):276–87.
25. Guo RJ, Eun RS, Lynch JP. The role of Cdx proteins in intestinal development and cancer. Cancer Biol Ther. 2004;3(7):593–601.
26. Pruenster M, Vogl T, Roth J, Sperandio M. S100A8/A9: From basic science to clinical application. Pharmacol Ther. 2016;167:120–31.
27. Reichman H, Moshkovits I, Itan M, Pasmanik-Chor M, Vogl T, Roth J, et al. Transcriptome profiling of mouse colonic eosinophils reveals a key role for eosinophils in the induction of s100a8 and s100a9 in mucosal healing. Sci Rep. 2017;7(1):7117.
28. Kim YS, Ho SB. Intestinal goblet cells and mucins in health and disease: Recent insights and progress. Curr Gastroenterol Rep. 2010;12(5):319–30.
29. Johansson MEV, Hansson GC. Immunological aspects of intestinal mucus and mucins. Nat Rev Immunol. 2016;16(10):639–49.
30. Cao Y, Raith MR, Griffith JF. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. Water Res. 2015;70:337–49.
31. Nshimyimana JP, Martin SL, Flood M, Verhougstraete MP, Hyndman DW, Rose JB. Regional variations of bovine and porcine fecal pollution as a function of landscape, nutrient, and hydrological factors. J Environ Qual. 2018;47(5):1024–32.
32. Campbell RK, Schulze K, Shaikh S, Ali H, Wu L, Raqib R, et al. Biomarkers of environmental enteric dysfunction among children in rural Bangladesh. Journal of Pediatric Gastroenterology and Nutrition. 2017;65(1):40–6.
33. Colston JM, Yori PP, Colantuoni E, Moulton LH, Ambikapathi R, Lee G, et al. A methodologic framework for modeling and assessing biomarkers of environmental enteropathy as predictors of growth in infants: An example from a Peruvian birth cohort. Am J Clin Nutr. 2017;106(1):245–55.
34. Arndt MB, Richardson BA, Mahfuz M, Ahmed T, Haque R, Gazi MA, et al. Plasma fibroblast growth factor 21 is associated with subsequent growth in a cohort of underweight children in Bangladesh. Curr Dev Nutr. 2019;3(5):1–10.

35. Arndt MB, Walson JL. Enteric infection and dysfunction—A new target for PLOS Neglected Tropical Diseases. PLoS Negl Trop Dis. 2018;12(12):1–6.
36. Anderson CJ, Kendall MM. Salmonella enterica serovar typhimurium strategies for host adaptation. Front Microbiol. 2017;8(OCT):1–16.
37. Jensen BH, Olsen KEP, Struve C, Krogfelt KA, Petersen AM. Epidemiology and clinical manifestations of enteroaggregative escherichia coli. Clin Microbiol Rev. 2014;27(3):614–30.
38. Barry R, Ruano-Gallego D, Radhakrishnan ST, Lovell S, Yu L, Kotik O, et al. Fecal neutrophil elastase-antiprotease balance reflects colitis severity. Mucosal Immunol. 2020;13(2):322–33.

CHAPTER III[2]

Understanding the Impact of Water and Sanitation on Child Health in Informal Settlements in Addis Ababa, Ethiopia.

3.1. Introduction

Recent work in urban areas highlights that new paradigms are necessary for urban water, sanitation, and hygiene (urban-WASH) interventions. In densely populated urban areas, overall fecal contamination is decoupled from traditional WASH indicators, suggesting that only transformational community wide interventions will be effective(1). An understanding of the interplay between utility provided water, sanitation technologies, and health is an important step in evaluating the type of transformational interventions that maybe necessary to improve urban-WASH. To this end, we conducted a WASH survey in informal settlements and evaluated the quality of utility provided water in Addis Ababa, Ethiopia.

The bulk of WASH studies have been conducted in rural areas, however given current rates of urbanization, 68% of the global population is expected to be urban by 2050, it is important to develop a better understanding of the kind of WASH interventions that are effective in dense, resource poor urban areas(2). In addition, it is not clear what lessons from rural WASH interventions can be transferred to urban settings. For example, in rural areas, a focus area has been to ensure that all households have access to a clean water source within a reasonable

[2] Chapter III is in the process of revision for publication.

distance and improved sanitation within a family compound(3). However, in urban areas, especially informal settlements, it may be more important to support public utilities to ensure that water of adequate quality is continuously provided, and central sewerage systems are built to safely dispose of excreta(3). However, central sewerage systems are not always an option, and data on their effectiveness is limited(3).

The goal of sanitation interventions is to eliminate or decrease pathogen input into the environment, depleting the environmental pathogen pool and decreasing the transmission potential. However, how to effectively implement sanitation technologies that interrupt transmission in heavily contaminated urban areas is still unclear. By their very nature, urban areas and especially informal settlements are characterized by extremely high population densities, resulting in a highly contaminated environment, which may be resistant to household level sanitation interventions if not done at scale. Informal settlements in urban areas therefore present a unique challenge, and traditional sanitation interventions may have to be adapted to the unique challenges that urban areas in low- and middle-income countries (LMICs) present.

Though the effectiveness of sanitation interventions in urban areas is unclear, the effectiveness of water interventions in reducing disease burden among urban populations has been known for more than a century(4). However, in the majority of cities in LMICs, water is supplied on an intermittent basis, with different parts of a city receiving water on a rotating basis(5). Intermittent water supply systems pose a higher disease risk because: 1) intermittent supplies are subject to increased microbial contamination by intrusion of outside matter during low pressure events; 2) microbial regrowth can occur in the distribution system during stagnant periods; 3) bacterial biofilms may scour off during repressurization; and 4) end users are forced to find alternative water sources during interruptions(6–9). The few studies that have looked at

the disease risks associated with intermittent water supply found that even short-term interruptions in systems with continuous supplies have been associated with diarrheal illness and in systems prone to interruption, longer water outages have been associated with an increased risk of diarrheal illness(10–14).

In addition, many water utilities in LMICs fail to properly disinfect water and maintain adequate disinfectant residual in the distribution system(15). The WHO recommends a chlorine residual of between 0.2-0.5mg/L be maintained in the distribution system, with this residual being especially important to ensure that there is no recontamination during the distribution or transport to the end user(15). An added challenge in maintaining chlorine residuals is that they generally decline with distance from the treatment plant with the rate of depletion affected by water flow velocity, residence time, pipe material and water pressure(15,16). However even with the maintenance of an adequate chlorine residual, the chlorine cannot protect from all pathogens, and can be over overcome by high concentrations of introduced contaminants(15,17). The maintenance of adequate residuals takes on an additional importance in intermittent systems where end users collect water for storage when water is available. Low chlorine residuals will drop below the recommended range during even short storage intervals, increasing the risk of water recontamination and bacterial growth in the water and storage containers(18,19). Another often over-looked shortcoming in many utilities is inadequate source water protection. Source water protection ensures lower pathogen loads in water taken in for treatment and ensures that pathogens resistant to disinfectant such as Cryptosporidium, are less likely make it into the distribution system.

Given the potential for failure at multiple points in the water distribution, implementing a multi-barrier approach (MBA) in the management of water distribution systems is crucial. The

MBA is an integrated multi-step system, with each step in the system designed to incrementally improve water quality. Steps in the MBA approach can be customized for the situation at hand and be implemented either centrally at a water utility or at the household level. The flexibility of the afforded by the MBA approach makes it especially useful for implementation in low resource settings where interventions must be customized to local conditions to ensure compliance and effectiveness.

From 2015 to 2019, we conducted a series of studies to evaluate the quality of utility water in Addis Ababa, Ethiopia. We conducted a WASH survey of 712 households in informal settlements to gauge 2-week diarrheal prevalence in infants aged 6-23 months and collected stool samples from a subset of infants . Water quality was gauged using indicator loads, chlorine residuals and high throughput sequencing. Infant stool samples were screened for a range of pathogens using droplet digital PCR (ddPCR). The aims of our study were to: 1) evaluate the quality of utility provided water in Addis Ababa using bacteriological indicators; 2) examine the how the distribution system characteristics impact bacterial communities; 3) examine the association between water quality, sanitation, and infant diarrheal prevalence; and 4) evaluate the association between water quality, sanitation, and pathogen carriage in infants in informal settlements.

3.2. Material and Methods

3.2.1. Study Site

Addis Ababa has an average elevation of 2,355m above sea level and has two rainy seasons: June to September and November to January. Addis Ababa was estimated to have a population of 3.2 million in 2015, accounting for 17% of Ethiopia's urban population(20). Only 10% of the people had access to a central sewerage system and 72% of residents did not have

access to adequate toilet facilities(20). It is however reported that 99% city inhabitants have access to a piped water source(20).

3.2.2. Water Sources and Treatment Plants

The city largely relies on surface and groundwater as the main water sources for drinking, domestic and industrial purposes. There are two surface water treatment plants, Legedadi and Gefersa, that together provided 52% of the city's treated water at the time this study was completed(21). The Legedadi water treatment plant was established in 1970 and is located 30 km northeast of the city center(21). It is the largest of the surface water treatment plants with a capacity of 195,000 m^3/day and delivered about 47% of the daily distributed drinking water for Addis Ababa in 2016(21). The Gefersa water treatment plant was established in 1940 and is located 20 km northwest of the city center(21). It has a capacity of 30,000 m^3/day and provided only 7% of the city's distributed drinking water in 2016(21). Both treatment plants use conventional treatment that includes pre-chlorination, coagulation, flocculation, sedimentation, sand filtration and post-chlorination, with a goal of maintaining an average chlorine residual of 0.8 mg/l in the distribution system(22).

3.2.3. Collection of Water Samples from the Distribution System

In 2015, a total of 38 water samples were collected from several locations from source to tap along both the Legedadi (n=22) and Gefersa (n=16) water distribution systems(22). For both the Gefersa and Legedadi distribution systems, samples were collected at the following points: 1) source water taken up by the treatment plants; 2) treated drinking water prior to being fed into distribution lines; 3) large, utility-maintained storage tanks in the distribution system that stored treated water ; 4) taps inside buildings; and 5) storage tanks at buildings(22). The number of samples by sampling point and water distribution system is provided in Table 3-1, along with

distances from the respective treatment plants. Water samples were collected in autoclave sterilized glass bottles supplemented with sodium thiosulfate to quench residual chlorine. Prior to sample collection water was run for 10 minutes to avoid capturing stagnant water(22). For culture-based analysis of water quality, 1L of water was collected, while for high throughput sequencing, 2L of water was collected(22).

3.2.4. Chemical and Microbiological Analysis of Distribution System Water Samples

Samples were field tested for temperature, pH, conductivity, and total dissolved solids (TDS) using a portable multiparameter meter or a portable pH meter (Oakton Instruments, Vernon Hills, IL USA). Free residual chlorine was measured in the field using the HACH DPD colorimetric method with a HACH DR890 colorimeter (Hach, Loveland, CO USA).

Microbial water quality was evaluated by measuring heterotrophic plate count (HPC), total coliform (TC) and fecal coliform (FC). All three were measured using the membrane filtration technique using and 0.45 µm gridded cellulose acetate membranes based on protocols described in Standard Methods(23–25). Samples for HPC analysis were diluted prior to filtration by adding 1mL of a drinking water sample to 99 mL of phosphate buffered solution and mixing by repeated inversion in sterile dilution bottles before vacuum filtering(22). The membrane was placed into a sterile petri dish containing R2A media and incubated at 28°C for five days(22). All cream-colored colonies were counted with a digital colony counter; results are given in units of CFU/ml(22). The enumeration of both TC and FC was done using three replicates of 100ml of undiluted sample vacuum filtered onto the gridded membranes(22). Membranes were transferred from the filtration rig using sterile forceps onto a sorbent pad saturated with membrane lauryl sulphate broth (MLSB)(22). Plates were incubated at 37°C for 4 hours to give cells time to recover from filtration(22). Plates TC counts were incubated at 37°C for 24 hours while, plates for FC counts

were incubated at 45°C for 24 hours(22). All yellow colonies were counted using a digital colony counter; results are given in units of CFU/100 mL(22).

3.2.5. DNA Extraction from Distribution System Water Samples and High Throughput Sequencing

Two liters of water per location were filtered through a polycarbonate membrane filter (0.22 µm EMD Millipore™ GTTP02500) and frozen immediately. All filtered samples were shipped frozen via overnight courier service to the Environmental Biotechnology Laboratory at the University of Michigan (UM), USA, and arrived frozen. Each membrane filter was cut into four equally sized pieces using a sterile knife and placed into a single vial to facilitate DNA extraction. Total genomic DNA was extracted from each vial using a Power Soil DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA) following the manufacturer's protocol and instructions. The extracted DNA was resuspended in a total volume of 50µl and prepared for high throughput sequencing. The concentration and purity of the extracted DNA was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and stored at -20°C until sequencing.

Extracted DNA was submitted for sequencing at the University of Michigan's sequencing core where Illumina MiSeq was performed with 2x250 paired-end chemistry. An amplicon library was generated for the V4 region of the 16S rRNA gene after two step amplification of the DNA fragment using universal dual index primers(26).

3.2.6. WASH Survey in Informal Settlements to Evaluate 2-Week Diarrheal Prevalence in Infants

In 2018, a WASH survey was conducted in 12 informal settlements spread north to south in Addis Ababa (Figure 3-1). Informal settlements for the survey were selected based on the

following criteria: 1) they had not been designated for redevelopment and residents were not being relocated to housing developments at the time of the study; and 2) the selected settlements would only be redeveloped in 2025, when Ethiopia is scheduled to achieve middle-income status.

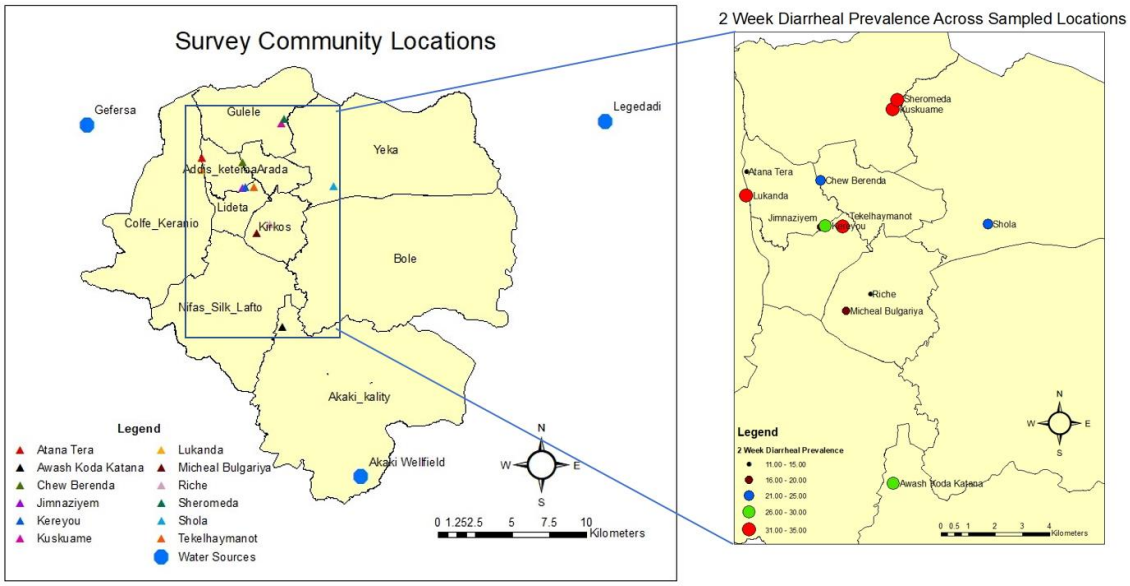


Figure 3-1: Locations of informal settlements and water treatment plants. Also shown is the 2-week diarrheal prevalence in each of the informal settlements.

Data on self-reported diarrheal disease, demographics, household assets, sanitation practices, food security, nutritional diversity, and anthropometry on a single index child in selected households was collected. Prior to the start of the main survey, health extension workers were recruited to carry out a census to identify households with infants aged 6-23 months. The survey was administered to 712 households located in the 12 informal settlements. Households from the census were then randomly selected for inclusion in the study. The number of households surveyed by location ranged from 50 to 74.

3.2.7. Infant Stool Sample Collection

136 infant fecal samples from the 12 study sites were collected. Stool samples were collected by giving selected households stool collection containers during the survey and requesting that they collect the index infant's stool over the course of the day. The collection containers were picked up in the late afternoon and stored at 4°C prior to processing the following day.

3.2.8. Nucleic Acid Extraction from Stool Samples

Nucleic acids were extracted using the ZymoBiomics DNA/RNA Mini Kit (Zymo Research, Irvine, CA). Standard kit protocol for nucleic acid isolation from fecal samples was used. Briefly, 200mg of stool was weighed out and placed in screw cap microcentrifuge tubes containing the DNA/RNA Shield Lysis Buffer. The tubes were packed in leak proof containers for transport to the University of Michigan where DNA and RNA were extracted as specified in the manufacturer's protocol within two months of sample collection. The quality and concentration of the extracted nucleic acids was measured using a Nanodrop Spectrophotometer (Thermo Scientific, Waltham, MA). Both DNA and RNA were used for pathogen identification.

3.2.9. Droplet Digital PCR

Pathogens screens used the QX200™ Droplet Digital™ PCR (ddPCR) system (Bio-Rad, Hercules, CA). DNA assays were setup by combining 10µl of ddPCR Supermix for Probes (no dUTP), primers at 900nM, probes at 250 nM and 60ng of template DNA or sterilized water for no template controls. RNA assays were setup using the One-Step RT-ddPCR Supermix (Bio-Rad, Hercules, CA) containing 5µl Supermix, 2µl Reverse Transcriptase, DTT at 15mM, primers at 900nM, probes at 250nM and 60ng of template RNA or 4µl of sterilized water for no template controls. Prior to setting up Rotavirus assays, samples were denatured at 95°C for 5 minutes and

kept on ice prior to adding the RNA to the reaction. Droplets were generated using the QX200™ AutoDG™ Droplet Digital™ PCR system (Bio-Rad, Hercules, CA). 40µl of the generated droplets were loaded into a 96-well plate and sealed using a PX1™ plate sealer (Bio-Rad, Hercules, CA). DNA thermocycling conditions consisted of enzyme activation at 95°C for 10 min; denaturation at 94°C for 30 s; annealing and extension at 58°C for 1 min; for 40 cycles, followed by enzyme deactivation at 98°C for 10 min, and a continuous hold at 4°C. RNA thermocycling conditions consisted of 60 min reverse transcription at 50°C, enzyme activation for 10 min at 95 °C, followed by 40 cycles of denaturation at 94°C for 30 s, annealing and extension at 58°C for 1 min; followed by enzyme deactivation at 98°C for 10 min, and a continuous hold at 4°C. All samples were run on the C1000 Touch™ thermocycler (Bio-Rad, Hercules, CA), with a lid temperature of 105°C, a sample volume of 40ul and a ramp rate of 2°C. On completion of the thermocycling, plates were read using the QX200™ Droplet Reader (Bio-Rad, Hercules, CA, USA) and QuantaSoft™ software (Bio-Rad, Hercules, CA). Samples were screened for eight bacteria (enteroaggregative *E.coli* (EAEC), enteropathogenic *E.coli* (EPEC), enterotoxigenic *E.coli* (ETEC), shigatoxigenic *E.coli* (STEC), *Shigella*, *Salmonella enterica*, *Campylobacter spp*, and *Enterococcus faecalis*), three protozoans (*Entamoeba histolytica*, *Giardia lamblia*, and *Cryptosporidium spp*), and two viruses (rotavirus and norovirus GI and GII). EAEC, EPEC and ETEC were each screened for two gene targets and the detection of either gene target was counted as a positive.

3.2.10. Droplet Digital PCR Data Processing

Samples were screened as described previously(27–29). All sample quantification was carried out using QuantaSoft™ software (Bio-Rad, Hercules, CA). Wells were checked and samples with <10,000 accepted droplets were rerun. To check the inter assay variability, 1/3 of

the samples were randomly selected to be re-run for both mRNA transcript and pathogen quantification, and coefficients of variation (CV) were calculated. If more than 5% of re-run samples had CVs higher than 15%, all assays were rerun.

3.2.11. Chlorine Decay Analysis of Distribution System Water

In 2019, residual chlorine measurements from standpipes and storage containers in two informal settlements were obtained using a Hach colorimeter. A total of 42 samples were collected. Standard protocol was followed except for not rinsing the vial with deionized water prior to each reading. Standpipes were not flushed, but instead a 100 mL graduated cylinder was filled and then used to wash out the vial three times before filling the vial for measurement.

Samples for testing chlorine decay rate constants were taken only from standpipes currently receiving water from the system. We sought to test at a location in pipes before they entered the community and then at the different endpoints of that branch. This was to test for a significant difference between the concentrations and decay rate constants along the distribution network.

Samples for K-tests were collected in 1L glass bottles. Prior to use, each bottle was thoroughly washed with soap and water and then left to dry. Upon collecting a sample for a K-test, the chlorine concentration was taken using the Hach colorimeter. Thereafter the chlorine concentrations using a Hach colorimeter were taken approximately every three hours except during the night. Samples were stored in the dark at room temperature. Chlorine residual values were then plotted against time with t=0 being the time of the first sample. The equation used to obtain the K value was:

$$Ln\left(\frac{C}{C_0}\right) = -k * t$$

Where:

C = chlorine concentrations (mg/L)

C_0 = initial chlorine concentration (mg/L)

k = chlorine decay rate constant (hrs^-1)

t = time (hrs)

Chlorine decay analysis was performed using the Microsoft Excel package.

3.2.12. Bioinformatic Analysis

Forward and reverse reads were processed using the DADA2 R-package(30,31). Forward and reverse reads were truncated at 240 and 200 nucleotides respectively. Paired-end reads were then merged, and chimeras removed. Taxonomy was assigned based on the Ribosomal Database Project (RDP) Training set (Version 16)(32). Further down downstream analysis was performed using the phyloseq package in R version 4.0.3(33). Beta-diversity was analyzed via PCoA using the Bray-Curtis dissimilarity index and the association between the water systems and sample collection points was evaluated using Adonis in the Vegan Package in R version 4.0.3(34). Alpha diversity was analyzed using the Shannon and Simpson indices(35–38).

3.2.13. Bacteriological Analysis

Descriptive statistics of selected variables are presented as either the mean and standard deviation or a percentage of the total. Mean colony counts and Wilcoxon Rank Sum Tests were used to assess differences in variables in the distribution water samples.

Outcome variables for the distribution water quality analysis were as follows: 1) log10 heterotrophic plate count, 2) log10 total coliform count, and 3) log10 fecal colony count. Exposure

variables included the water treatment plant supplying the water, collection point, free chlorine levels, temperature, TDS, pH, and distance from the treatment plant.

3.2.14. Survey Analysis

The two outcomes for survey analysis were two- week diarrheal prevalence and pathogen diversity. Diarrhea was defined as three or more loose, watery, or bloody stools in a 24-hour period, while pathogen diversity was defined as the sum of unique pathogen types detected in a single stool sample.

Two-week diarrheal prevalence was treated as a binary variable and was evaluated using a multi-level logistic regression model with informal settlements included as a random effect. The primary exposure variables were household water source, water treatment type and household sanitation type. Additional covariates were included based on prior knowledge of their association with diarrheal prevalence and were further evaluated using bivariate analysis between each variable the primary outcome before final inclusion in the model. Socioeconomic status (SES) variables included the level of education of the primary caregiver, household income, dwelling ownership, and residency status. Infant age, water intermittency, and household income were standardized prior to inclusion in the model.

A Poisson model with robust error variance was used the asses the association between pathogen diversity and all outcome variables while controlling for the confounders. We looked at pathogen diversity as measure of overall pathogen exposure. The final model was arrived as described above.

All statistical analysis was done using R version 4.0.3. The multilevel logistic model was fit using the lme4 package and the Sandwich package was used to obtain robust standard errors for the Poisson model(39–41).

3.2.15. Human Subjects Approval

Study protocols were approved by institutional review boards at the University of Michigan (HUM00115103), and the Addis Ababa University (IRB/029/2017). Parents or legal guardians gave verbal, informed consent prior to participation or collection of data.

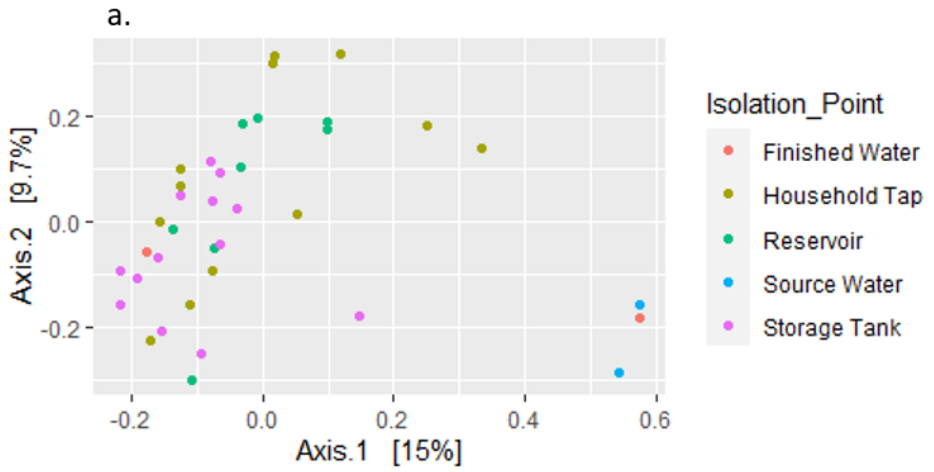
3.3. Results

3.3.1. Bacteriological and Physical Analysis of Water Quality by Distribution System

A total of 38 water samples were collected from source to tap along both the Legadadi (N=22) and Gefersa (N=16) water distribution systems. The samples comprised two source water samples at each of the plants and two samples of finished water before the water entered the distribution system, 12 water samples taken at taps, 13 water samples from household tanks, and nine water samples from utility storage reservoirs. HPCs were significantly different between the Legedadi and Gefersa systems (7453.33 cfu/ml and 16450 cfu/ml respectively, p-value=0.03). Similarly, the total of number of coliforms was significantly different between the Legedadi and Gefersa distribution systems (16.33 cfu/100ml and 101.33 cfu/100ml respectively, p-value=0.008). Distances to taps were also significantly different between the Legedadi and Gefersa systems (21.53km and 13.50km respectively, p-value=0.03). Household storage tanks were also on average closer to the treatment plant in the Gefersa system compared to the Legedadi system. Summary statistics for all variables are presented in Table 3-1.

Linear regression models were used to examine the association between HPC, total coliform counts and fecal coliform and distribution system water (Table 3-2). Compared to the Gefersa system, the Legedadi system had significantly lower HPC, TC counts, and FC counts (β=-1.10, p-value <0.001, β=-2.41, p-value=0.001, and β=1-34, p-value=0.03 respectively). Total dissolved solids were also positively associated with HPC (β=0.01, p-value=0.04) as was distance (β=0.06, p-value=0.01). Compared to reservoirs, taps had lower FC counts (β=-0.95, p-value=0.05).

There was a significant difference in the bacterial communities by sampling point (Figure 3-2a) (PERMANOVA R^2= 0.16, p=0.03). Three of four finished and source water samples clustered together, and overall, 25% of the variation was explained by the two axes. There were no significant differences in bacterial community by water distribution system (Figure 3-2b) (PERMANOVA R^2 = 0.04, p= 0.21).



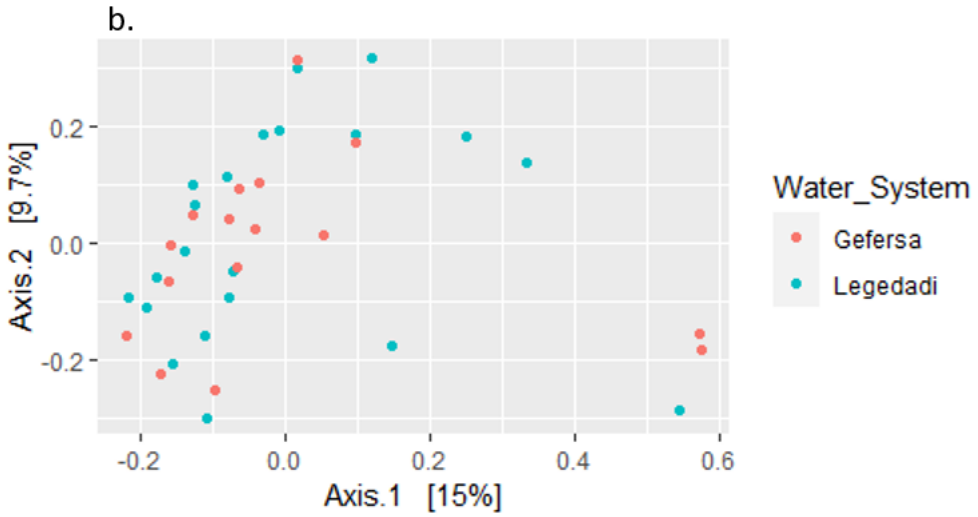


Figure 3-2: Relationship between the water bacterial communities in water distribution systems and collection points. NMDS plots of Bray-Curtis dissimilarity comparing a) communities by different water distribution systems, and b) by collection point. Axes represent the variation explained by each axis based on eigen values .

While there was a decrease in Shannon index values down the distribution system, there was little variation in Simpson index values across the different sampling points (Figure 3-3).

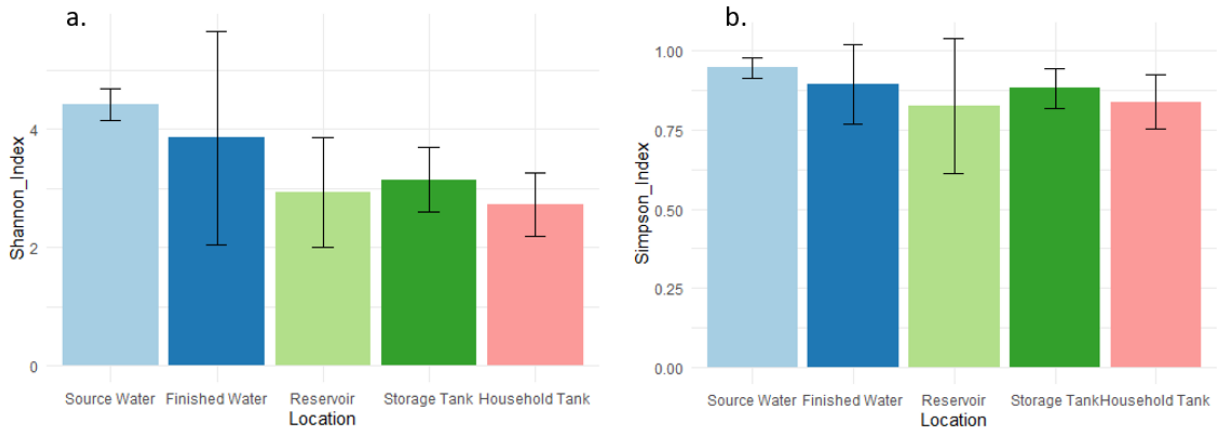


Figure 3-3: Diversity index values by sampling location: a) Shannon index and b) Simpson index. There were no significant differences in index value between any of the sampling points.

There was an increase in the relative abundance of *Gammaproteobacteria* across the distribution system (Figure 3-4). *Gammaproteobacteria* were least abundant in source and finished water samples (22.91% and 19.84% respectively). In samples from reservoirs,

household storage tanks, and households taps, *Gammaproteobacteria* predominated (49.85%, 51.75%, and 49.24% respectively). However, the differences in the relative abundance of *Gammaproteobacteria* by sampling point were not significant.

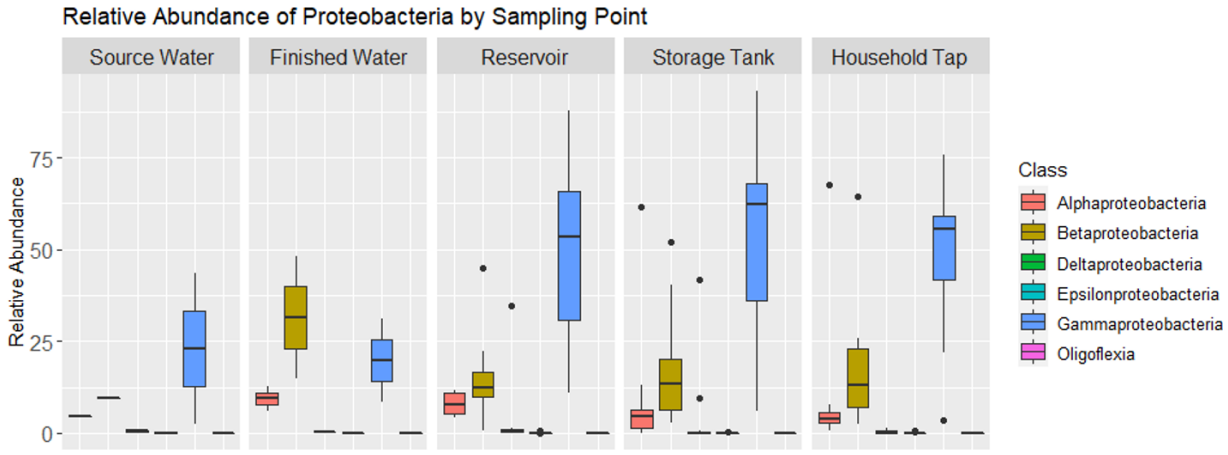


Figure 3-4: Relative abundance of *Proteobacteria* across the water distribution system in Addis Ababa, Ethiopia. There is an increase in abundance on *Gammaproteobacteria* down the distribution system with *Gammaproteobacteria* predominating in reservoirs, storage tanks and household taps.

3.3.2. 2-week Diarrheal Prevalence in Infants Aged 6-23 Months

Households with infants aged 6-23 months were surveyed in 12 informal settlements (Table 3-3). The average age of infants was 14.21 months, and our sample was evenly split between male and female infants (53.19 % of the infants were female). The majority households obtained water from a piped source in their yard (75.36%) and reported not having water for an average 3.70 days during the week. Most households did not treat their water (56.67%), but of those who did treat their water prior to use, most used bleach (27.74% of households reported bleaching their water prior to use). Most households had access to latrine with a slab (50.87%) or to a latrine without a slab (34.34%).

A multilevel logistic regression model was used to examine the association between household WASH variables and 2-week diarrheal prevalence in infants aged 6-23 months (Table

3-4). Compared to households that had water pumped into their dwellings, infants in households that obtained water from pipes in their yards and water from public taps had significantly lower odds of diarrhea (OR= 0.35, 95% CI 0.16, 0.76 and OR=0.39, 95% CI 0.15, 1.00 respectively). Similarly, compared to households who did not treat their water, infants in households that boiled their water and infants in households that filtered their water also had significantly lower odds of diarrhea (OR= 0.40, 95% CI 0.19, 0.86 and OR=0.23, 95% CI 0.06, 0.84 respectively). Water intermittency and the presence of soap at a handwash station were also associated with two-week diarrheal prevalence (OR= 1.13, 95% CI 0.99, 1.39 and OR=0.24, 95% CI 0.08, 0.67 respectively). In addition, compared to infants in households with an earthen floor, infants in households with carpeted floors had a significantly higher odds of diarrhea (OR= 4.95, 95% CI 1.88, 13.04).

3.3.3. Infant Stool Pathogen Types and Water, Sanitation and Hygiene Variables.

A Poisson regression model was used to evaluate the association between infant stool pathogen diversity and household water, sanitation, and hygiene variables (Table 3-5). Compared to the Gefersa water system, infants in households that obtained their household water from the Legedadi system had a significantly lower risk of testing positive for a pathogen (RR=0.82, 95% CI 0.71, 0.95). Similarly, compared to households that did not treat their household drinking water, infants in households that filtered their water had a lower risk of testing positive for a pathogen (RR=0.57, 95% CI 0.37, 0.85).

3.3.4. Measurement of Disinfectant Residuals and Disinfectant Decay

Water samples collected from two informal settlements, both part of the Gefersa distribution system, had chlorine residual levels measured and chlorine decay constants calculated (Table 3-6). There was a non-significant difference in chlorine residual levels between Atana Tera and Lukanda (0.79 mg/L and 0.50 mg/L respectively). There was also a marked but non-significant

difference in chlorine residuals between pipes and storage tanks at the two locations. In Atana Tera, pipes had average chlorine residual level of 1.03 mg/L, compared to 0.60 mg/L in storage tanks. Similarly, in Lukanda, pipes had an average chlorine disinfectant level of 0.65 mg/L, compared to 0.39 mg/L in storage tanks. We also noted spatial variation in disinfectant residuals. (Figure 3-5).

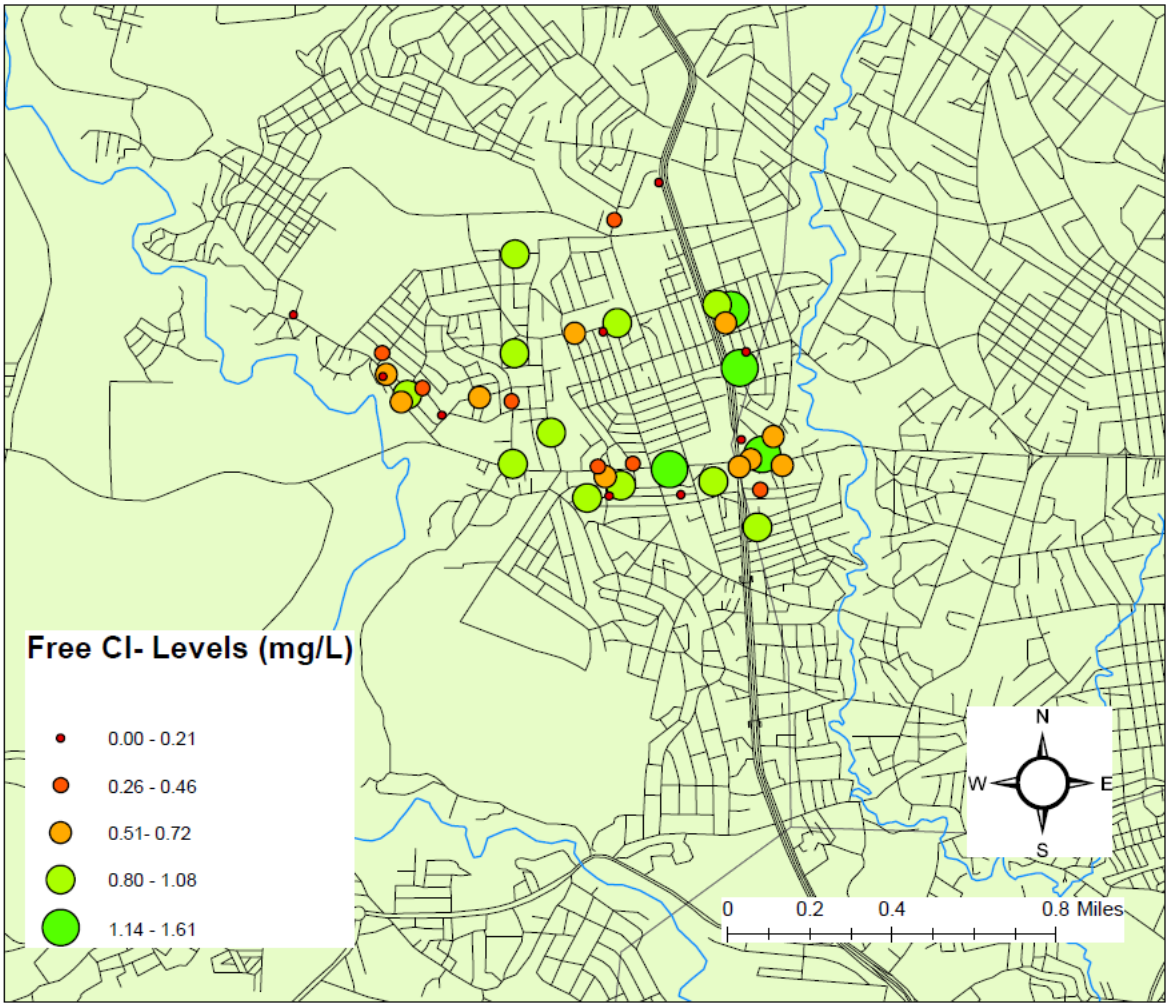


Figure 3-5: Map of spatial variation in free chlorine disinfectant levels in the communities of Atana Tera and Lukanda.

On average, chlorine levels dropped below the threshold of 0.2mg/L in stored water in less than 24 hours (Figure 3-6a). At the maximum observed decay rate, chlorine residuals dropped below 0.2 mg/L in just over 10 hours (Figure 3-6b).

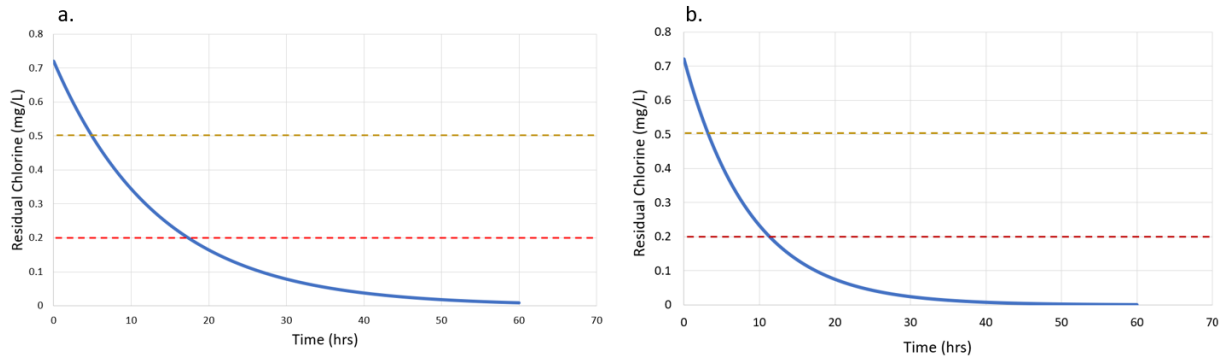


Figure 3-6: Disinfectant decay rates in water stored in plastic containers in Addis Ababa, Ethiopia. a) the average decay rate in plastic containers, b) the maximum observed disinfectant decay rate. The dotted lines define the safe range for disinfectant residual (0.2-0.5mg/L) recommended by the WHO.

3.4. Discussion

Urban-WASH requires a holistic approach that accounts for WASH indicators at both the household and city level. Assessing utility provided water using bacteriological coupled with community surveys can help understand the inter-play between WASH indicators at different levels and their impact on disease. Water quality in Addis Ababa as measured by HPCs, TCs, and FCs was significantly different by treatment plant in Addis Ababa (Table 3-2). There were significant differences in bacterial community structure by sampling point (Figure 3-2a). There was also a change in bacterial community composition down the water distribution system, with the abundance of *Gammaproteobacteria* increasing down the distribution system (Figure 3-4). At the household level, infants in households that obtained their water from taps in their yards or from public taps had lower odds of diarrhea than infants in households with premise plumbing (Table 3-4). Water treatment also resulted in lower odds of diarrhea, with infants in households that either boiled or filtered water having lower odds of diarrhea compared to households that did not treat their water (Table 3-4). The risk of testing of positive for a pathogen also differed between water distribution systems and the risk of testing positive for a pathogen was also

significantly lower in infants from households that filtered their water compared to those that did not treat their water (Table 3-5). We also observed variation in free chlorine levels over small spatial distances (Figure 3-5) and a rapid decay of chlorine residuals in stored utility supplied water (Figure 3-6). Our findings point to utility water being a potential driver of diarrheal disease in informal settlements.

The bacteriological contamination that we detected in Addis Ababa utility water is likely the result of failures in both the treatment of source water and of contamination within the water distribution system. Our findings reinforce previous findings on the quality of municipal water in Addis Ababa(42). The significant difference in bacteriological indicators by water treatment plant is evidence that there were significant differences in the bacteriological quality of water pumped into the distribution system by treatment plant at the time of sampling. In addition, our finding that the relative abundance *Gammaproteobacteria* increases down the distribution system is of particular concern given their potential as reservoirs of antibiotic resistance or as opportunistic pathogens. Among the possible explanations for the differences in bacterial community composition by sampling point and the increase in the relative abundance of *Gammaproteobacteria* down the distribution system are : 1) unstable chlorine residuals in the distribution system may cause a shift in the microbial community structure down the distribution system; and 2) changes in the microbial community are driven by changes in the relative contribution of biofilms to the community, with contributions varying by chlorine concentrations(43). It is likely that the changes in microbial community structure that we observed are driven unstable chlorine residuals impacting the makeup and contribution of biofilms to the microbial community in the distribution system. In a study from China looking at disinfectant effects on biofilm bacterial communities, Mi et al. (2015) found that high chlorine

disinfectant levels decreased the abundance *Gammaproteobacteria* while increasing the proportion of *Betaproteobacteria* in biofilms(44). They also observed that *Gammaproteobacteria* predominated in undisinfected biofilms and biofilms exposed to low or medium chlorine doses(44). This is similar to what we observed, with Betaproteobacteria dominating the community in water samples from the treatment plants, where chlorine residuals are likely to be more constant, while *Gammaproteobacteria* increase in abundance down the distribution, where chlorine residuals are more variable. The dominance of *Gammaproteobacteria* down the distribution system could also explain why see a reduction in the odds of diarrhea for infants in households who do not use premise plumbing as their main water source. Bacterial communities in premise plumbing in Addis Ababa may be dominated by opportunistic pathogens, and premise plumbing is unlikely to be flashed as thoroughly as public taps are through higher use volume.

Our finding that boiling and filtration, rather than the use of bleach, are most effective in reducing the odds of diarrhea in infants also point to other potential deficiencies in utility provided water in Addis Ababa. Utility water may be contaminated with chlorine resistant pathogens such as Cryptosporidium. The presence of such pathogens in the system could be indicative of failures upstream in the treatment process and possibly of highly contaminated source water at both plants. The failure of bleach to reduce the odds of diarrhea could however also be due to deficiencies in the way that chlorine is utilized at the household level. Chlorine treatment could be less effective at the household level for the following reasons: 1) long water storage periods (median ~3 days, in our sample) increase the likelihood of in-home contamination; and 2) the improper use and incorrect dosing of chlorine given the water quality and baseline contamination levels already present in utility water, would further lower the

effectiveness of chlorine treatement(18). Further studies are needed to evaluate the effectiveness of in-home chlorine use in Addis Ababa.

Our findings point to potential multiple failure points in the treatment and distribution of water in Addis Ababa. Given the systemic shortcomings in the distribution system, multiple treatment processes need to be in place. Without adequate water treatment at the household level, the water consumed has only gone through only one 'barrier' at the treatment plant, which as the data suggests is likely inadequate. Safe water management in Addis Ababa should therefore include effective household level treatment steps until the utility water is brought up to quality. Furthermore, water quality should be assessed at multiple points to identify where key deficiencies occur. An MBA approach would allow for the adoption of more effective and sustainable mitigation measures be it at the household or city level. Any mitigation measures taken should also be coupled with regular surveillance of vulnerable populations to accurately gauge the effectiveness of interventions in a timely manner and to identify next steps.

We also found that carpeted floors resulted in significantly higher odds of diarrhea compared to earthen floors, pointing to the existence other of indirect pathogen transmission pathways in highly contaminated settings . In a setting where sanitation sharing is the norm, carpeted floors may act as trap for pathogens from multiple households, resulting in much higher household pathogen loads than other floor types and serve as an indirect route of pathogen transmission to children(45,46). This trapping of contamination by carpets is further compounded by the fact that carpets are harder to keep clean in low resource settings, resulting in a buildup of pathogens in the carpet material. The presence of a household pathogen pool in carpets may especially increase the risk of infection with more persistent pathogens as infants start to explore their environment.

Our results also highlight the nuances of personal hygiene practices in highly contaminated urban settings on disease risk. Household sanitation type did not have any effect on the odds of diarrhea in infants, but the presence of soap at a handwash station significantly reduced the odds of diarrhea in infants. The effectiveness of handwashing with soap in reducing diarrhea is well documented, but it's importance in highly contaminated urban settings takes on additional weight(47–50). The effectiveness of personal hygiene interventions may be due to their ability to disrupt direct and ubiquitous pathogen transmission pathways, transmission pathways that are present regardless of sanitation access or environmental contamination levels. On the other hand, interventions that are focused on the availability and access to sanitation technologies often disrupt indirect transmission pathways. Though effective in the long term, the disruption of indirect transmission pathways may require that a technology be deployed on a large scale and with community buy-in and high compliance, possibly explaining the recent unexpected results from large scale intervention trials on the effectiveness of sanitation interventions(47,51,52). The thresholds needed for disruption of multiple indirect transmission pathways may be especially high in dense urban areas given the proximity of individuals and extremely high shedding rates of some pathogens.

Our study also highlights some of the challenges of using joint monitoring program (JMP) service ladders for evaluating water and sanitation in informal settlements. More than 80% of the households surveyed reported having a water source either in their yards or in their premises, placing them in them in the safely managed rung of the service ladder if water were available when needed and was free of fecal and chemical contamination. However, as our data shows, water quality in Addis Ababa is highly variable, with disinfectant residuals varying over small spatial distances and bacteriological indicators differing significantly by water treatment

plant. In addition, our finding that premise plumbing increases the odds of diarrhea compared to public and yard taps indicates that premise plumbing may only be beneficial in systems with continuous flow and without bacteriological contamination. The evaluation urban water systems should therefore be geared toward understanding where failures in the distribution system occur. Measuring water quality only at the end user as it the norm makes it hard to establish causal associations between water quality deficiencies and upstream failure points given multiple possible failure points in distribution systems. Furthermore, defining the risk associated with different failures in urban water systems is key. A water supply system that is intermittent, but able to supply water with the correct disinfectant residuals and bacteriological indicators may be preferable to a less intermittent system that is unable to provide water of adequate quality. Studies evaluating the risks associated with different types of failures in intermittent urban water systems are lacking and should be prioritized to inform on the most cost-effective mitigation measures that low-resource water utilities should prioritize(53,54). The human dimension of intermittent water supplies also must be appreciated. In urban areas such as Addis Ababa, end-users face systemic failures resulting in the complete unavailability of water for several days, causing them to change water use behavior and switch to alternative water sources that may potentially be less safe. The switching between water sources results in oscillating levels of disease risk, with disease risk in households potentially driven by the source that is less safe. Capturing the oscillating levels of risk would require longitudinal study designs that would specifically take into the account the switching between sources.

Understanding the impact of water distribution systems on disease risk in LMICs is vital for the development of effective WASH interventions for urban areas. In our analysis, we found a significant difference in bacteriological indicators by treatment plant, indicating water

treatment deficiencies. We also show that in-home water treatment, either by boiling or by filtration, is key to reducing the odds of diarrhea, highlighting utility provided water as potential driver of disease among infants. More studies are needed to characterize the impact of both water and sanitation on health in informal settlements and the disease risk associated with each of them.

Table 3-1: Descriptive statistics of water quality by distribution system. Wilcoxon-Rank tests were used to check for differences in between the Legedadi and Gefersa water distribution systems. Significant differences are bolded.

Variable		N	Legedadi	Gefersa	p-value
HPC (CFU/mL)	**Tap (SD)**	12	**7458.33 (4179.27)**	**16450 (4700.39)**	**0.03**
	Tank (SD)	13	9083.33 (3509.78)	13483.33 (7575.59)	0.40
	Reservoir (SD)	9	5923.81 (1905.62)	7866.67 (1697.06)	0.38
Total Coliforms (CFU/100mL)	**Tap (SD)**	12	**16.33 (10.87)**	**101.33 (61.23)**	**0.008**
	Tank (SD)	13	27.50 (23.50)	57.96 (54.75)	0.22
	Reservoir (SD)	9	21.86 (11.53)	40.83 (13.43	0.19
Fecal Coliforms (CFU/100mL)	Tap (SD)	12	11.33 (6.93)	17.67 (2.37)	0.20
	Tank (SD)	13	15.40 (9.60)	17.54 (9.39)	0.56
	Reservoir (SD)	9	14.24 (6.77)	19.500 (8.72)	0.43
Free Chlorine (mg/L)	Tap (SD)	12	0.51 (0.19)	0.50 (0.14)	1
	Tank (SD)	13	0.44 (0.15)	0.51 (0.21)	0.76
	Reservoir (SD)	9	0.63 (0.18)	0.65 (0.07)	1
Temperature (^{0}C)	Tap (SD)	12	21.25 (0.90)	21.25 (0.5)	0.93
	Tank (SD)	13	21.20 (0.84)	20.88 (0.64)	0.47
	Reservoir (SD)	9	20.71 (1.11)	20.50 (0.71)	1
EC (µS/cm)	Tap (SD)	12	603.50 (43.56)	600.5 (29.64)	0.80
	Tank (SD)	13	624.8 (76.27)	565 (64.34)	0.12
	Reservoir (SD)	9	608.57 (44.39)	621.00 (29.70)	0.46
TDS (mg/L)	Tap (SD)	12	301.75 (21.78)	300.35 (14.81)	0.80
	Tank (SD)	13	312.40 (38.13)	285.00 (25.98)	0.12
	Reservoir (SD)	9	304.29 (22.19)	310.50 (14.85)	0.46
pH	Tap (SD)	12	7.35 (0.37)	7.24 (0.29)	0.80
	Tank (SD)	13	6.96 (0.30)	7.07 (0.49)	0.82
	Reservoir (SD)	9	7.11 (0.45)	7.51 (0.71)	0.30
Distance from Plant (km)	**Tap (SD)**	**12**	**21.53 (3.32)**	**13.50 (4.60)**	**0.03**
	Tank (SD)	**13**	**21.10 (2.19)**	**12.36 (4.09)**	**0.01**
	Reservoir (SD)	9	22.03 (3.55)	8.5 (0.71)	0.05

Table 3-2: Associations between bacteriological indicators and distribution system water. Significant associations are bolded.

Variable	Log Heterotrophic Plate Count			Log Total Coliform Count			Log Fecal Coliform Count		
	Estimate	**SE**	**p-value**	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**
Treatment Plants									
Gefarsa		REF			REF			REF	
Legedadi	**-1.10**	**0.26**	**<0.001**	**-2.41**	**0.66**	**0.001**	**-1.34**	**0.58**	**0.03**
Collection Point									
Reservoir		REF			REF			REF	
Tank	0.40	0.21	0.06	-0.55	0.52	0.31	-0.52	0.46	0.27
Tap	0.32	0.20	0.13	-0.93	0.52	0.09	**-0.95**	**0.46**	**0.05**
Free Cl (mg/L)	-0.16	0.44	0.72	-1.98	1.13	0.09	-1.42	1.00	0.17
Temp (^{0}C)	-0.08	0.11	0.45	0.50	0.27	0.07	0.39	0.24	0.11
TDS (mg/L)	**0.01**	**0.00**	**0.04**	0.01	0.01	0.16	0.00	0.01	0.53
pH	0.10	0.19	0.61	0.23	0.49	0.64	-0.01	0.43	0.99
Distance (km)	**0.06**	**0.02**	**0.01**	0.10	0.06	0.10	0.07	0.05	0.17

Table 3-3: Summary statistics of selected demographic, household, and sanitation variables (N=712).

Variable	Summary Statistics
Infant Age (SD)	14.21 (5.06)
Female (%)	53.19
Caregiver Education Level (%)	
No Education	17.24
Primary	50.87
Secondary	21.74
Vocational and Higher	10.14
Household Income (Ethiopian Birr) (SD)	2069.76 (1833.82)
Floor Material (%)	
Earth Sand or Dung	48.70
Wood	2.02
Cement or Ceramic Tiles	45.94
Carpet	3.33
Household Water Source (%)	
Piped into Dwelling	6.23
Piped into Yard	75.36
Public Tap	16.52
Other	1.36
Days without Water in a Week (SD)	3.70 (1.88)
Water Treatment Practices (%)	
No Treatment	56.67
Boil	10.00
Bleach	27.54
Filtration	5.65
Household Sanitation (%)	
Flush	6.39
VIP	7.10
Latrine with Slab	50.87
Latrine without Slab	34.34
No Sanitation	4.35

Table 3-4: Associations between two-week diarrheal prevalence in infants and water, sanitation, and hygiene variables. Significant associations are bolded.

	Unadjusted Estimate			Adjusted Estimate		
Variable	**OR**	**95% CI**		**OR**	**95% CI**	
Water Plant						
Gefersa		REF			REF	
Legedadi	0.95	0.61	1.49	0.89	0.51	1.53
Akaki	0.96	0.46	1.99	0.85	0.30	2.37
Household Water Source						
Piped into Dwelling		REF			REF	
Piped into Yard	**0.37**	**0.19**	**0.74**	**0.35**	**0.16**	**0.76**
Public Tap	**0.40**	**0.18**	**0.90**	**0.39**	**0.15**	**1.00**
Other*	1.50	0.41	5.56	1.68	0.39	7.27
Water Treatment						
No Water Treatment		REF			REF	
Boil	**0.46**	**0.22**	**0.96**	**0.40**	**0.19**	**0.86**
Bleach	0.89	0.58	1.36	0.96	0.61	1.51
Filtration**	**0.25**	**0.07**	**0.87**	**0.23**	**0.06**	**0.84**
Days with No Water	1.13	0.94	1.36	**1.13**	**0.99**	**1.39**
Household Water Storage Container						
Do Not Store		REF			REF	
Jerrycan	1.06	0.67	1.68	1.13	0.69	1.84
Plastic Bucket	0.92	0.49	1.73	0.97	0.49	1.93
Plastic and Iron Bucket	1.03	0.54	1.96	0.87	0.43	1.77
Other	2.05	0.60	7.04	1.99	0.53	7.41
Household Sanitation						
Flushing Toilets		REF			REF	
VIP	0.84	0.26	2.76	0.92	0.24	3.48
Latrine with Slab	0.80	0.29	2.21	0.78	0.24	2.54
Latrine without Slab	0.59	0.21	1.67	0.62	0.19	2.06
No Sanitation	1.28	0.37	4.48	1.44	0.35	5.93
Soap at Handwash Station						
No		REF			REF	
Yes	**0.30**	**0.11**	**0.80**	**0.24**	**0.08**	**0.67**
Feces on Toilet Floor						
No		REF			REF	
Yes	0.92	0.53	1.58	0.88	0.49	1.57
Floor Material						
Earth, Sand or Dung		REF			REF	
Wood	1.46	0.37	5.78	1.71	0.41	7.23
Cement or Ceramic	1.47	0.97	2.23	1.51	0.96	2.38
Carpet	**4.28**	**1.74**	**10.52**	**4.95**	**1.88**	**13.04**

Table 3-5: Associations between infant stool pathogen diversity and water, sanitation, and hygiene variables. Significant associations are bolded.

Variable	Unadjusted Estimate			Adjusted Estimate		
	Risk Ratio	95% CI		Risk Ratio	95% CI	
Water Plant						
Gefersa	REF			REF		
Legedadi	**0.85**	**0.74**	**0.99**	**0.82**	**0.71**	**0.95**
Akaki	0.91	0.68	1.23	0.80	0.56	1.15
Household Water Source						
Piped into Dwelling	REF			REF		
Piped into Yard	1.10	0.86	1.42	0.98	0.74	1.31
Public Tap	1.30	0.97	1.74	1.14	0.82	1.60
Bottled Water	1.08	0.53	2.16	0.98	0.49	1.96
Water Treatment						
No Treatment	REF			REF		
Boil	1.03	0.80	1.32	0.97	0.77	1.22
Bleach	1.09	0.95	1.26	1.04	0.89	1.22
Filter*	**0.58**	**0.40**	**0.84**	**0.57**	**0.37**	**0.85**
Days with No Water	**0.91**	**0.85**	**0.97**	0.92	0.85	1.10
Household Water Storage Container						
Do Not Store	REF			REF		
Jerrycan	0.97	0.81	1.15	1.01	0.84	1.21
Plastic Bucket	1.11	0.85	1.45	1.16	0.87	1.56
Plastic and Iron Bucket	1.06	0.87	1.29	1.19	0.94	1.50
Household Sanitation						
Flush or VIP	REF			REF		
Latrine with Slab	1.08	0.82	1.43	1.11	0.87	1.42
Latrine without Slab	1.05	0.79	1.40	1.13	0.86	1.49
No Sanitation	1.23	0.78	1.93	1.44	0.85	2.12
Soap at Handwashing Station						
No	REF			REF		
Yes	0.86	0.60	1.25	0.99	0.69	1.43
Feces on Toilet Floor						
No	REF			REF		
Yes	1.04	0.87	1.25	0.98	0.82	1.18
Floor Material						
Earth, Sand or Wood	REF			REF		
Cement or Ceramic	0.90	0.77	1.04	0.92	0.78	1.08
Carpet	1.14	0.80	1.61	1.19	0.83	1.72

Table 3-6: Disinfectant residual levels in pipes and storage tanks at two informal settlements; Atana Tera and Lukanda.

Atana Tera			
Variable	**Pipe (N=6)**	**Storage Tank (N=8)**	**p-value**
Free Cl- (mg/L) (SD)	1.03 (0.53)	0.60 (0.30)	0.07

Lukanda			
Variable	**Pipe (N=8)**	**Storage Tank (N=10)**	**p-value**
Free Cl-(mg/L) (SD)	0.65 (0.33)	0.39 (0.24)	0.09

References

1. Holcomb DA, Knee J, Sumner T, Adriano Z, de Bruijn E, Nalá R, et al. Human fecal contamination of water, soil, and surfaces in households sharing poor-quality sanitation facilities in Maputo, Mozambique. Int J Hyg Environ Health . 2020;226:113496.
2. United Nations U. World Urbanization Prospects. Vol. 12. 2018.
3. WHO/UNICEF. Progress on household drinking water, sanitation, and hygiene 2000-2017.
4. Woods RI, Watterson PA, Woodward JH. The causes of rapid infant mortality decline in England and Wales, 1861-1921. Part II. Popul Stud (NY). 1989;43(1):113–32.
5. Danilenko A, van den Berg C, Macheve B, Moffitt LJ. The IBNET Water Supply and Sanitation Blue Book 2014. 2015.
6. Kumpel E, Nelson KL. Intermittent Water Supply: Prevalence, Practice, and Microbial Water Quality. Environ Sci Technol. 2016;50(2):542–53.
7. Kumpel E, Nelson KL. Mechanisms affecting water quality in an intermittent piped water supply. Environ Sci Technol. 2014;48(5):2766–75.
8. Coelho ST, James S, Sunna N, Abu Jaish A, Chatiia J. Controlling water quality in intermittent supply systems. Water Sci Technol Water Supply. 2003;3(1–2):119–25.
9. Bivins AW, Sumner T, Kumpel E, Howard G, Cumming O, Ross I, et al. Estimating Infection Risks and the Global Burden of Diarrheal Disease Attributable to Intermittent Water Supply Using QMRA. Environ Sci Technol. 2017;51:7542–51.
10. Yassin MM, Amr SSA, Al-Najar HM. Assessment of microbiological water quality and its relation to human health in Gaza Governorate, Gaza Strip. Public Health. 2006;120(12):1177–87.
11. Nygård K, Wahl E, Krogh T, Tveit OA, Bøhleng E, Tverdal A, et al. Breaks and maintenance work in the water distribution systems and gastrointestinal illness: A cohort study. Int J Epidemiol. 2007;36(4):873–80.
12. Hunter PR, Chalmers RM, Hughes S, Syed Q. Self-reported diarrhea in a control group: a strong association with reporting of low-pressure events in tap water. Clin Infect Dis. 2005;40(4):32–4.
13. Huang LY, Wang YC, Liu CM, Wu TN, Chou CH, Sung FC, et al. Water outage increases the risk of gastroenteritis and eyes and skin diseases. BMC Public Health. 2011;11.
14. Abu Amr SS, Yassin MM. Microbial contamination of the drinking water distribution system and its impact on human health in Khan Yunis Governorate, Gaza Strip: Seven years of monitoring (2000-2006). Public Health. 2008;122(11):1275–83.
15. Lee EJ, Schwab KJ. Deficiencies in drinking water distribution systems in developing countries. J Water Health. 2005;3(2):109–27.
16. Egorov A, Ford T, Tereschenko A, Drizhd N, Segedevich I, Fourman V. Deterioration of drinking water quality in the distribution system and gastrointestinal morbidity in a Russian city. Int J Environ Health Res. 2002;12(3):221–33.
17. Gadgil A. Drinking Water in Developing Countries. Annual Rev Energy Environ. 1998;23(1):253–86.
18. Levy K, Anderson L, Robb KA, Cevallos W, Trueba G, Eisenberg JNS. Household effectiveness vs. laboratory efficacy of point-of-use chlorination. Water Res. 2014;54:69–77.
19. McLaughlin LA, Levy K, Beck NK, Shin G-A, Meschke JS, Eisenberg JN. An Observational Studyn on the Effectiveness of Point-Of-Use Chlorination. International Perspect. 2009;71(8):48–53.
20. United Nations Human Settlements Programme (UN-Habitat). The state of Addis Ababa. 2017.

21. Addis Ababa Water and Sewerage Authority.
22. Atanfu B, Desta A, Assefa F. Microbial community structure and diversity in drinking water supply, distribution systems as well as household point of use site in Addis Ababa city. Pre-Print. 2021.
23. 9215 HETEROTROPHIC PLATE COUNT. Standard Methods for the Examination of Water and Wastewater. Stand Methods.
24. 9225 DIFFERENTIATION OF COLIFORM BACTERIA. Standard Methods for the Examination of Water and Wastewater. Stand Methods.
25. 9222 MEMBRANE FILTER TECHNIQUE FOR MEMBERS OF THE COLIFORM GROUP. Standard Methods for the Examination of Water and Wastewater. Stand Methods.
26. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. Appl Environ Microbiol. 2013;79(17):5112–20.
27. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Anal Chem. 2012;84:1003–11.
28. Cao Y, Raith MR, Griffith JF. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. Water Res. 2015;70:337–49.
29. Nshimyimana JP, Martin SL, Flood M, Verhougstraete MP, Hyndman DW, Rose JB. Regional Variations of Bovine and Porcine Fecal Pollution as a Function of Landscape, Nutrient, and Hydrological Factors. J Environ Qual. 2018;47(5):1024–32.
30. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):4–5.
31. Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. Bioconductor workflow for microbiome data analysis: From raw reads to community analyses. F1000Research. 2016;5:1–49.
32. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37(SUPPL. 1):141–5.
33. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4).
34. Oksanen J, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M. The vegan package. Community Ecol Packag. 2007;10:631–7.
35. Simpson EH. Measurement of diversity. Nature. 1949;688(1943):688.
36. Magurran A. Measuring Biologcial Diversity. Blackwell Publishing. 2004. p. 256.
37. Lemos LN, Fulthorpe RR, Triplett EW, Roesch LFW. Rethinking microbial diversity analysis in the high throughput sequencing era. J Microbiol Methods. 2011;86(1):42–51.
38. Kim BR, Shin J, Guevarra RB, Lee JH, Kim DW, Seol KH, et al. Deciphering diversity indices for a better understanding of microbial communities. J Microbiol Biotechnol. 2017;27(12):2089–93.
39. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1).
40. Zeileis A. Object-oriented computation of Sandwich estimators. J Stat Softw. 2006;16(9):1–11.
41. Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. J Stat Softw. 2004;11(10):128–9.

42. Wolde AM, Jemal K, Woldearegay GM, Tullu KD. Quality and safety of municipal drinking water in Addis Ababa City, Ethiopia. Environ Health Prev Med. 2020;25(1):1–6.
43. Stanish LF, Hull NM, Robertson CE, Kirk Harris J, Stevens MJ, Spear JR, et al. Factors influencing bacterial diversity and community composition in municipal drinking waters in the Ohio River basin, USA. PLoS One. 2016;11(6):1–21.
44. Mi Z, Dai Y, Xie S, Chen C, Zhang X. Impact of disinfection on drinking water biofilm bacterial community. J Environ Sci (China). 2015;37:200–5.
45. Budge S, Barnett M, Hutchings P, Parker A, Tyrrel S, Hassard F, et al. Risk factors and transmission pathways associated with infant Campylobacter spp. prevalence and malnutrition: A formative study in rural Ethiopia. PLoS One. 2020;15(5):1–17.
46. Exum NG, Olortegui MP, Yori PP, Davis MF, Heaney CD, Kosek M, et al. Floors and toilets: Association of floors and sanitation practices with fecal contamination in Peruvian Amazon peri-urban households. Environ Sci Technol. 2017;176(10):139–48.
47. Luby SP, Agboatwalla M, Painter J, Altaf A, Billhimer WL, Hoekstra RM. Effect of intensive handwashing promotion on childhood diarrhea in high-risk communities in Pakistan: A randomized controlled trial. J Am Med Assoc. 2004;291(21):2547–54.
48. Hirai M, Roess A, Huang C, Graham JP. Exploring the link between handwashing proxy measures and child diarrhea in 25 countries in sub-Saharan Africa: A cross-sectional study. J Water Sanit Hyg Dev. 2017;7(2):312–22.
49. Kamm KB, Feikin DR, Bigogo GM, Aol G, Audi A, Cohen AL, et al. Associations between presence of handwashing stations and soap in the home and diarrhoea and respiratory illness, in children less than five years old in rural western Kenya. Trop Med Int Heal. 2014;19(4):398–406.
50. Ejemot-Nwadiaro RI, Ehiri JE, Arikopo D, Meremikwu MM, Critchley JA. Handwashing promotion for preventing diarrhea. Cochrane Database Syst Rev. 2015;(9).
51. Humphrey JH, Mbuya MNN, Ntozini R, Moulton LH, Stoltzfus RJ, Tavengwa N V, et al. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. Lancet Glob Heal. 2019;7(1):e132–47.
52. Null C, Stewart CP, Pickering AJ, Dentz HN, Arnold BF, Arnold CD, et al. Effects of water quality , sanitation , handwashing , and nutritional interventions on diarrhoea and child growth in rural Kenya : a cluster-randomised controlled trial. Lancet Glob Heal. 2018;6(3):E316–29.
53. Ercumen A, Arnold BF, Kumpel E, Burt Z, Ray I, Nelson K, et al. Upgrading a Piped Water Supply from Intermittent to Continuous Delivery and Association with Waterborne Illness: A Matched Cohort Study in Urban India. PLoS Med. 2015;12(10):1–24.
54. Ercumen A, Gruber JS, Colford JM. Water distribution system deficiencies and gastrointestinal illness: A systematic review and meta-analysis. Environ Health Perspect. 2014;122(7):651–60.

CHAPTER IV[3]

Determinants of Pathogen Contamination of the Environment in the Greater Yangon Area, Myanmar

4.1. Introduction

Implementing safely managed sanitation has been a trusted Water and Sanitation Health (WASH) intervention to eliminate or decrease pathogen input into the environment, decreasing transmission potential and disease burden. However, recent WASH trial results have called into question the effectiveness of sanitation interventions and its impact on pathogen transmission (1–3), leading many to argue for the need to improve our understanding of the role that environment plays in enteropathogen transmission. Estimating contamination levels provides a more direct measure of exposure than traditional assessments of whether a sanitation technology is in place. These measurements also provide for an objective comparison of multiple environmental transmission pathways. Some of these pathways are affected by traditional sanitation interventions, while others are not.

In this cross-sectional study, we tested environmental samples for a range of bacteria, protozoa, helminths, and viruses from three urban sites in Yangon, Myanmar and two rural sites on the outskirts of Yangon to assess environmental pathogen loads. The aim of this study was to

[3] Chapter IV has been submitted for publication with the following co-authors: Ther Aung, Khin Khin Han, Pam Jagger and Joseph N.S Eisenberg.

understand how environmental characteristics drive contamination and how this differs in rural versus urban settings. We add to a growing number of studies in low and middle-income countries that take a similar approach to studying environmental pathogen loads and their association with environmental factors (4,5).

Diarrhea is the fourth-leading cause of death in children under five in Myanmar, with deaths from diarrheal disease higher in urban than in rural areas (11.3% vs. 6.8% respectively) (6,7). Hospital-based surveillance studies in Yangon have identified rotavirus as a major cause of diarrhea in children; an analysis of five years of hospital surveillance data identified rotavirus in 42-56% of hospitalized acute gastroenteritis cases each year (8,9). Yangon itself represents the dynamics of environmental exposure in a setting of rapid urbanization and limited investment in WASH infrastructure(10). For example, in a rapidly expanding peri-urban neighborhood in Yangon it was that found that 36% of surveyed households disposed of household waste into nearby water bodies, such as stream and ponds (11). Similarly, a geo-spatial analysis of water-borne diseases in Myanmar found that Yangon had the highest number of water-borne disease cases and outbreaks from 1991-2018 due to a combination of high population density and limited WASH infrastructure (12).

Pathogen contamination has traditionally been measured using indicator organisms such as enterococcal bacteria and fecal *E.coli,* which are often limited in their ability to generalize across all relevant pathogens (13–15). Specifically, these organisms are poorly correlated with other infectious organisms (4). Instead, a selection of host specific pathogens may provide a more accurate picture of environmental contamination. Given that we now have affordable technology to measure pathogens in the environment, this study will help us better understand if environmental pathogen transmission pathways are similar between different pathogen types and

what role the environment may have in the transmission of different pathogens in urban and rural locations.

The dynamics of pathogen concentrations in the environment are driven by shedding rates into the reservoir, pathogen decay, and environmental factors such as heavy rain events (16). The specific dynamic processes will differ by 1) pathogen, i.e., shedding rates from infectious individuals into the environment, persistence, and transport processes are all pathogen dependent; 2) environmental factors such as local climate and hydrological processes; 3) social factors that affect how people interact with the environment; and 4) geographic location, e.g., rural vs. urban (17–19).

Environmental enteropathogen risks have been more extensively studied in rural compared to urban settings. Rural settings are typically characterized by lower population densities, resulting in pathogen shedding that is more dispersed and stochastic. Contamination levels in urban settings, on the other hand, tend to be uniformly high, regardless of local sanitary conditions and other traditional indicators of risk such as socioeconomic status, suggesting the need for community level transformational changes (20). With 68% of the of global population projected to live in urban areas by 2050, it is important that we improve our understanding of the role of the environment in pathogen transmission within densely populated urban areas (21).

4.2. Material and Methods

4.2.1. Study Sites

This study was conducted in the Yangon Region in lower Myanmar, bordered to the south by the Gulf of Martaban. The Yangon Region has a population of approximately 7,360,730 (22), with two thirds of the population living in the urban area. The Yangon Region is

administratively divided into 45 townships, of which 33 fall under the jurisdiction of Yangon city, the largest city in Myanmar.

The study sites were in two townships in the northern part of Yangon, Hlaingtharya and Htantabin (Figure 4-1). Hlaingtharya is located within the Yangon city administrative boundary. It is the most populous township in the city with 686,827 inhabitants and population density of 8,229 persons/km (23). Htantabin borders Hlaingtharya to the north and sits outside of the city's administrative boundary. Htantabin is predominantly rural with a population of 145,792 and a population density of 240.2 person/km (24).

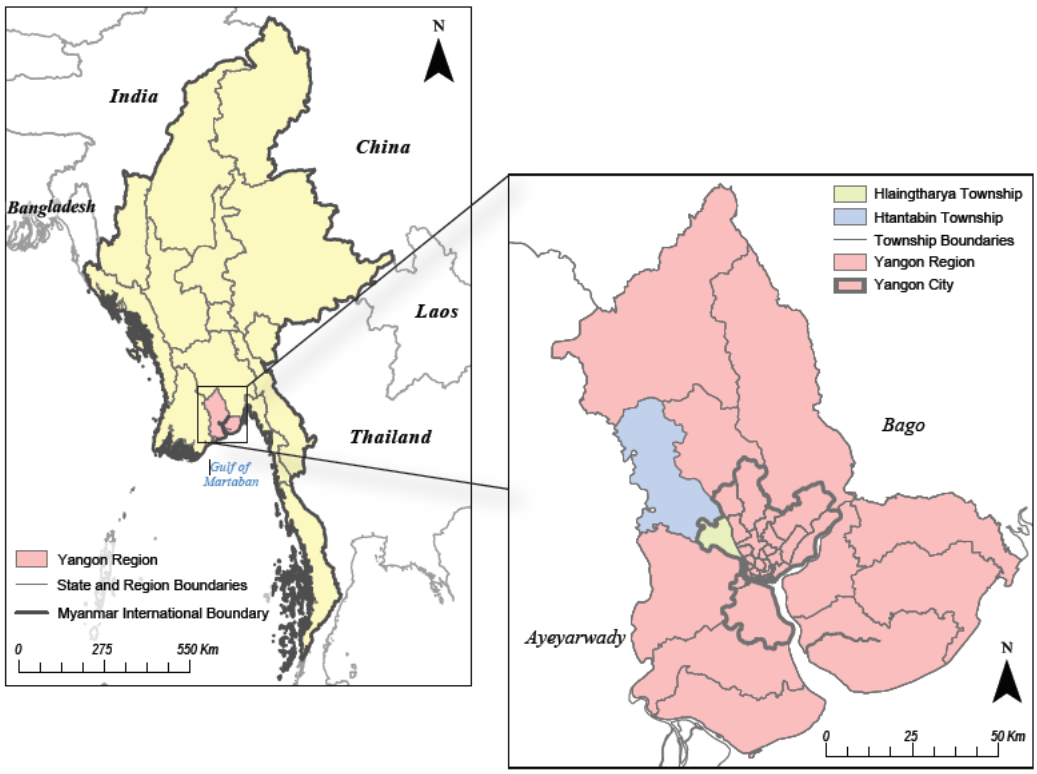


Figure 4-1: Map of our study sites in the greater Yangon area. Our three urban sites were located in the Hlaingtharya Township in the north of Yangon, while out two rural sites were located in the Hlantabin Township outside of city boundaries. Map Provided by Dr. Pamela Jagger.

We purposively selected three urban wards in Hlaingtharya: Shwe Lin Ban (SLB); Ward 7; and Yeokkan, and two rural sites in Htantabin Township. These sites represent heterogeneity

in broad categories of living conditions including density, permanence of structures, infrastructure, and public service provision.

4.2.2. Environmental Site Characteristics

Yangon City has an average elevation of about 23 meters above sea level. Yangon has a tropical monsoon climate under the Koppen Climate Classification (25). The rainy season in Yangon runs from May through October, with a dry season from November to April. Total annual precipitation averages 2783 mm and with average highs ranging from 29°C to 36°C and average lows ranging from 18°C to 25°C. Sample collection was conducted over an eight-day period at the beginning of October, at the end of the rainy season. No rain events occurred over the eight days of sample collection.

4.2.3. Sample Collection Point Selection

Sample collection points were selected to capture the variability in environmental characteristics across the study areas. Maps were used to define sampling areas and random latitude and longitude pairs were generated using the website geomidpoint.com. The generated coordinate pairs were then entered into Google Maps app and were ground truth to ensure that selected sampling points were accessible and were feasible for sampling. If a coordinate pair was at a location not viable for sampling, a sample point was selected within a roughly 50m radius of the original coordinates. Final sampling points were recorded using a Garmin Etrex 10 meter and were then entered into Google Maps. Field staff used mobile phones and the Google Maps app to navigate to the marked coordinates. On average, we estimate that field staff were able to arrive within at least 2 meters of the selected sampling points.

At each site, a sampling point was either standing water, soil, or a man-made drainage ditch. We collected a water sample from a drainage ditch if there was standing water. We

collected a soil samples from a drainage ditch if there was no standing water and the soil was moist at the time of sampling, indicating recent water flow in the conduit (Figure 4-2).

a. b.

c. d.

Figure 4-2: Selected images of study site. Images a and b are taken in urban informal settlements with a. showing a drainage ditch on the outside of a dwelling and b. showing a dwelling raised above standing flood water. c and d are taken at the rural study site. In both images the thatched structure is a dwelling.

4.2.4. Sample Collection and Site Observation

Prior to sample collection, field staff were trained to ensure standardized hygienic sampling protocols. Soil was collected using an alcohol sterilized metal tablespoon inserted into the ground to the depth of the spoon bowl. The soil sample was then transferred to a sterile WhirlPak bag (Sigma-Aldrich, St Louis, MO). Surface water was collected by skimming a sterile 50ml sterile propylene tube over the water's surface. All collected samples were then stored on ice packs in a cooler and transported to a field laboratory at the West Yangon University for processing within 4-6 hours.

While collecting samples, field staff also recorded the following details at sample collection points:

- Whether or not the samples were collected from a man-made conduit for water, regardless of whether it contained water at the time of sampling. This information was used to create the dichotomous variable **Drainage Ditch**, i.e., whether the ditch was man-made**.**
- Whether animals were observed within 50m of the sampling point. These observations were used to create the dichotomous variable **Presence of Animals.**
- Whether a sanitation facility was observed within 50m of the sampling point. These observations were used to create the dichotomous variable **Nearby Toilet.**
- Whether any animal feces visible within 5m of the sampling point. These observations were used to create the dichotomous variable **Presence of Animal Feces.**
- Whether uncollected garbage was present within 50m of the sampling point. These observations were used to create the dichotomous variable **Uncollected Garbage.**

In addition, data on site elevation and the state of the observed sanitation facilities were also collected.

4.2.5. DNA and RNA Extraction

DNA and RNA were extracted from both soil and water using the ZymoBIOMICS DNA/RNA MiniKit (Zymo Research, Irvine, CA). For the extraction of nucleic acid from soil

samples, 250mg of soil from a WhirlPak bag was weighed out and transferred to a DNA/RNA Shield Lysis tube. Water samples were vacuum filtered through 0.45µm nitrocellulose filters which were then transferred to DNA/RNA Shield Lysis tubes. All sample tubes were stored in a cool dry area before transport back to the University of Michigan for extraction. Samples were processed within 3 weeks of collection as par the manufacturers protocol. The quality and concentration of the extracted nucleic acids was measured using a Nanodrop Spectrophotometer (Thermo Scientific, Waltham, MA).

4.2.6. Droplet Digital PCR

Sample pathogen loads were quantified using the QX200™ Droplet Digital™ PCR (ddPCR) system (Bio-Rad, Hercules, CA). Prior to setting up PCR assays, all environmental assays were diluted 1:10 to ensure that there was no inhibition(30). DNA assays were setup by combining 10µl of ddPCR Supermix for Probes (no dUTP), primers at 900nM, probes at 250 nM and 4µl of template DNA or sterilized water for no template controls. RNA assays were setup using the One-Step RT-ddPCR Supermix (Bio-Rad, Hercules, CA) containing 5µl Supermix, 2µl Reverse Transcriptase, DTT at 15mM, primers at 900nM, probes at 250nM and 4µl of template RNA or 4µl of sterilized water for no template controls. Prior to setting up Rotavirus assays, samples were denatured at 95°C for 5 minutes and kept on ice prior to adding the RNA to the reaction. Droplets were generated using the QX200™ AutoDG™ Droplet Digital™ PCR system (Bio-Rad, Hercules, CA). 40µl of the generated droplets were loaded into a 96-well plate and sealed using a PX1™ plate sealer (Bio-Rad, Hercules, CA). DNA thermocycling conditions consisted of enzyme activation at 95°C for 10 min; denaturation at 94°C for 30 s; annealing and extension at 58°C for 1 min; for 40 cycles, followed by enzyme deactivation at 98°C for 10 min, and a continuous hold at 4°C. RNA thermocycling conditions consisted of 60 min reverse

transcription at 50°C, enzyme activation for 10 min at 95°C, followed by 40 cycles of denaturation at 94°C for 30 s, annealing and extension at 58°C for 1 min; followed by enzyme deactivation at 98°C for 10 min, and a continuous hold at 4°C. All samples were run on the C1000 Touch™ thermocycler (Bio-Rad, Hercules, CA), with a lid temperature of 105°C, a sample volume of 40µl and a ramp rate of 2°C. All samples were run in duplicate with repeats run on different days. On completion of the thermocycling, plates were read using the QX200™ Droplet Reader (Bio-Rad, Hercules, CA, USA) and QuantaSoft™ software (Bio-Rad, Hercules, CA). Samples were screened for eight bacteria (enteroaggregative *E.coli* (EAEC), enteropathogenic *E.coli* (EPEC), enterotoxigenic *E.coli* (ETEC), shigatoxigenic *E.coli* (STEC), *Shigella*, *Salmonella enterica*, *Campylobacter spp*, and *Enterococcus faecalis*), three protozoans (*Entamoeba histolytica*, *Giardia lamblia*, and *Cryptosporidium spp*), three helminths (*Ancylstoma duodenale*, *Ascaris trichuris*, *Stronyloides sterocalis*), and two viruses (rotavirus and norovirus GI and GII). EAEC, EPEC and ETEC were each screened for two gene targets and the detection of either gene target was counted as a positive.

4.2.7. Sample Quantification

Samples were quantified as described previously(26–28). Briefly, QuantaSoft™ software (Bio-Rad, Hercules, CA) was used to check all wells, and samples with <10,000 accepted droplets were excluded from analysis. The threshold for differentiating negative from positive droplets was determined by setting it one standard deviation above the negative droplets on the no template controls(27,28). Following the setting of the threshold, all wells were visually inspected to ensure that the thresholding was correct. Wells with less than three positive droplets were considered negative and assigned a value equal to the limit of detection (LOD) divided by 2 (LOD/2)(29). All samples where one of the duplicates failed to generate > 10,000 droplets or

where duplicates had a coefficient of variation greater > 10% were re-run. Final concentrations of gene copies per gram of soil or per millimeter of surface water were obtained by multiplying by appropriate dilution factors.

4.2.8. Statistical Analysis

Descriptive statistics of selected variables are presented as either the mean and standard deviation or a percentage of the total. Mean gene counts and Wilcoxon Rank Sum Tests were used to assess differences in gene loads by substrate between urban and rural areas. Non-detect samples were represented as the LOD in Wilcoxon Rank Sum tests.

Outcome variables were derived from gene loads measured using ddPCR. Outcomes were as follows: 1) log10 *E.faecalis* gene count, 2) log10 pathogenic gene count, 3) log10 bacterial gene count, 4) log10 viral gene count, and 5) pathogen diversity, defined as the sum of unique pathogen types detected in a single sample. For statistical analysis, non-detect samples had LOD/2 values applied to them.

Exposure variables include elevation in meters, and variables to indicate a drainage ditch sample, the presence of domestic animals, sanitation facilities, animal feces, and uncollected garbage. All models also included two potential confounders: substrate type (soil versus water) and collection location (rural versus urban).

Left censored regression models controlling for exposure variables and confounders were used to assess the association between the first four outcome variables and a specific exposure. A Poisson model with robust error variance was used to assess the association between pathogen diversity and all outcome variables while controlling for the confounders. All analysis was done using R version 4.0.3. Censored regression models were fit using the VGAM package and the Sandwich package was used to obtain robust standard errors for the Poisson model (30,31).

4.2.9. Ethical Approval

This study was part of a larger study implemented by the University of North Carolina at Chapel Hill (UNC-Chapel Hill), which was conducted following a protocol approved by its Institutional Review Board (#18-2735). This study did not involve any human or animal subjects nor were any personal identification data from the larger study used.

4.3. Results

4.3.1. Study Site Characteristics

Samples were collected from 117 points, comprising 79 water samples and 38 soil samples. 95 of those points were from three urban sites and 22 from two rural sites. All sampling sites were lower in elevation than the average for the Yangon area (23m), with rural areas lower in elevation than urban areas (5.00m and 7.06m respectively). We observed a number of differences comparing the urban and rural sites, including fewer domestic animals (57.6% vs. 95.7%), fewer visible toilet facilities (18.2% vs. 52.2%), and more visible uncollected garbage (21.2% vs. 0.0%) in urban compared to the rural sites (Table 4-1). Although urban sites had fewer visible toilets, they were more likely to be in better condition (5.6% vs. 41.7% toilets were observed to be dilapidated).

4.3.2. Pathogen Frequency, Loads and Diversity by Sampling Location

We collected 60 water samples from three urban sites, 19 water samples from two rural sites, 35 soil samples from the three urban sites, and three soil samples from one rural site. *Enterococci* were detected in 115 (98.3%) of the sampled points (Figure 4-3). At least one pathogen was detected in 73.7% of all water samples and in 55.3% of all soil samples. The most

common pathogen in both water and soil was rotavirus, which was detected in 88.6% and 54.3% of samples, respectively.

Bacterial pathogens were isolated more often in urban vs. rural water samples. For example, EAEC, EPEC and ETEC were detected in 66.7%, 55.0%, and 46.7% of urban water samples respectively, whereas in rural water samples, the only detected bacterial pathogens were ETEC and *S.enterica*, each found in 5.3% of the samples. Rotavirus was the most common non-bacterial pathogen in both urban and rural water (91.7% and 79.0% of the samples respectively). Additionally, water samples in urban sites had a much richer pathogen pool dominated by bacterial pathogens, compared to rural sites that had only five pathogens detected: two bacteria, one helminth, and two viruses (Figure 4-3).

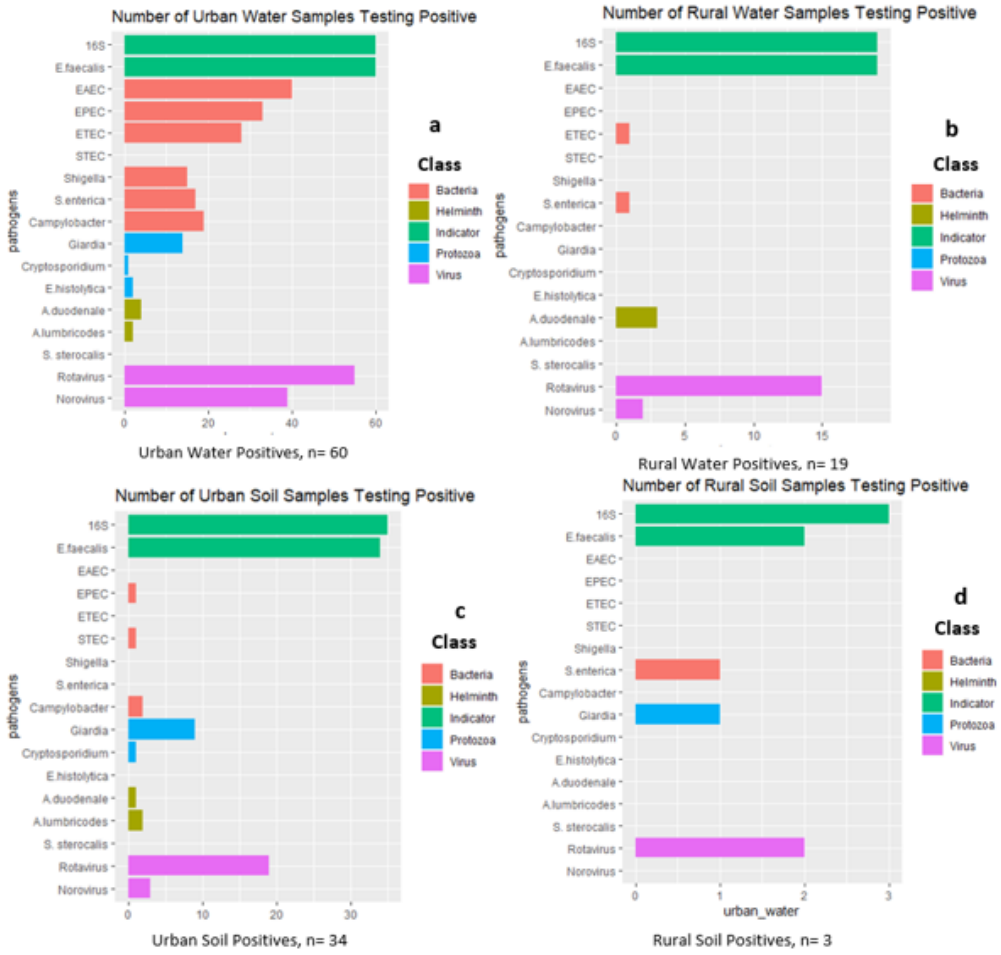
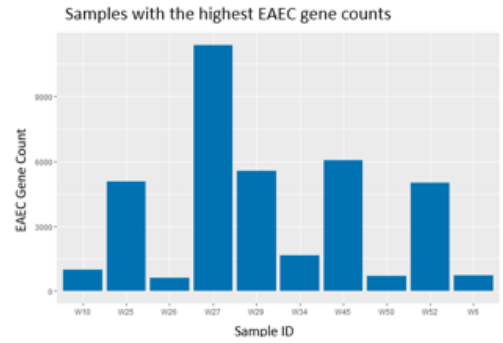
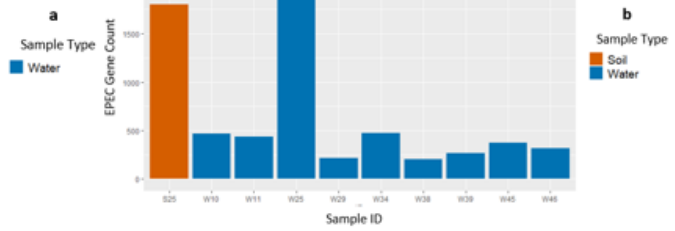


Figure 4-3 The number of samples testing positive for indicators, 16S DNA or pathogenic gene markers in urban water samples (a), rural water samples (b), urban soil samples (c), and rural soil samples (d) from sampling sites in the greater Yangon area.

Urban soil samples contained fewer pathogens than urban water samples. For example, *Campylobacter*, the most frequently detected bacterial pathogen in urban soil, was found in only 5.7% of the samples. In both urban and rural soil samples, rotavirus and *Giardia* were detected in a substantial number of samples (54.3% and 66.7% for rotavirus and 25.7% and 33.3% for *Giardia* respectively). As with water samples, urban soil samples had higher pathogen diversity than rural soil samples.

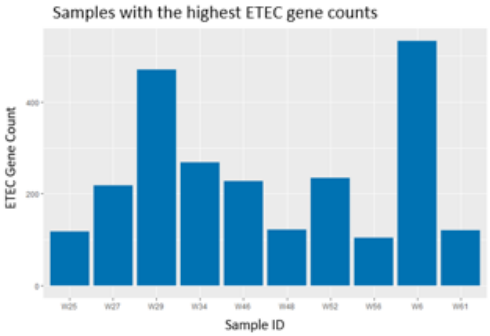
Gene counts, a measure of the quantity of marker genes, varied by both by pathogen and substrate (water vs. soil) (Figure 4-4). The highest *E. coli* pathotype and *Shigella* gene counts were generally detected in water samples. In general, EAEC gene counts in water were higher than both EPEC and *Shigella* gene counts, which in turn was higher than ETEC counts. In contrast, *Campylobacter* and *Giardia* gene counts were highest in soil samples. Norovirus gene counts were appreciably lower than rotavirus gene counts. Rotavirus gene counts were highly variable, with the highest rotavirus loads collected in two soil samples (1108556.12 and 28968.51 genes/g respectively), and counts in water samples being orders of magnitude lower, ranging from 3601.38-1411.00 genes/ml. The highest norovirus loads were detected in two soil samples (1302.34 genes/g and 1090.08 genes/g respectively).

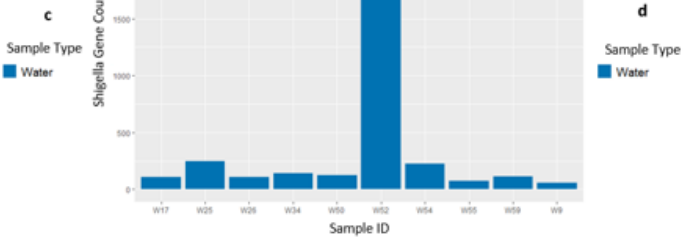

Samples with the highest EAEC gene counts
EAEC Gene Count
Sample ID


a
Sample Type
Water

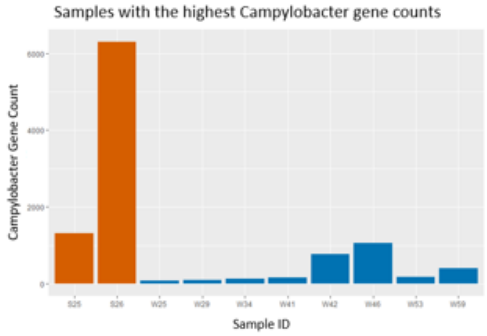
Samples with the highest EPEC gene counts
EPEC Gene Count
Sample ID

b
Sample Type
Soil
Water


Samples with the highest ETEC gene counts
ETEC Gene Count
Sample ID

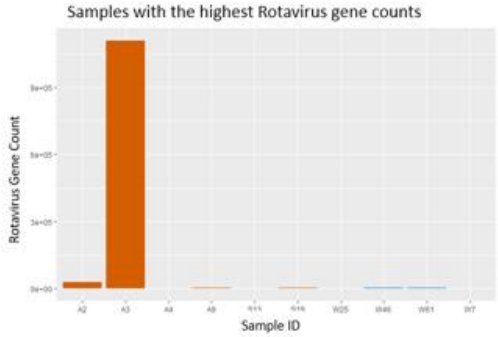

c
Sample Type
Water
d
Sample Type
Water

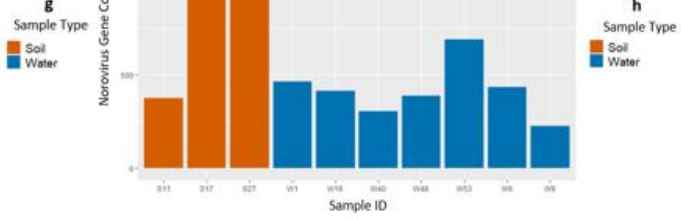
Samples with the highest Shigella gene counts
Shigella Gene Count
Sample ID


Samples with the highest Campylobacter gene counts
Campylobacter Gene Count
Sample ID


e
Sample Type
Soil
Water
f
Sample Type
Soil
Water

Samples with the highest Giardia gene counts
Giardia Gene Count
Sample ID


Samples with the highest Rotavirus gene counts
Rotavirus Gene Count
Sample ID


g
Sample Type
Soil
Water
h
Sample Type
Soil
Water

Samples with the highest Norovirus gene counts
Norovirus Gene Count
Sample ID

Figure 4-4: Top ten samples with the highest gene counts of a) EAEC, b) EPEC, c)ETEC, d) *Shigella*, e) *Campylobacter*, f) *Giardia*, g) Rotavirus, and h) Norovirus.

In general, pathogen loads per sample in both water and soil were significantly higher in urban areas compared to rural areas (Table 4-2). In urban water samples, bacterial pathogens had the highest geometric mean gene counts (4.31E+02, 95% CI 3.16E+02, 5.88E+02) followed by viruses (1.40E+02, 95% CI 9.50E+01, 2.03E+02). In rural water samples, bacteria also had the highest geometric mean gene counts (1.41E+02, 95% CI 1.39E+02, 1.42E+01). In urban soil samples, bacteria had the highest geometric mean gene counts (1.96E+02, 95%CI 1.33E+02, 2.90E+02), followed by viruses (1.20E+02 95%CI, 4.28E+01, 3.35E+02).

4.3.3. Association between Gene Counts and Site Characteristics

Censored regression models were used to examine the association between collection point characteristics and gene counts (Table 4-3). Samples collected from drainage ditches were significantly associated with log10 *Enterococci* and log10 pathogenic bacteria gene counts (RR=3.02, 95% CI 1.17, 7.77 and RR=2.31, 95% CI 1.43, 3.72 respectively). Uncollected garbage was also significantly associated with log10 *Enterococci* and log10 bacterial gene counts (RR=3.23, 95% CI 1.07, 9.72 and RR=1.98, 95% CI 1.13, 3.45 respectively). Elevation was only associated with log10 bacterial gene counts (RR=0.94, 95% CI 0.88, 0.99).

4.3.4. Association between Pathogen Diversity and Site Characteristics

A Poisson model was used to evaluate the association between sample pathogen diversity and site characteristics (Table 4-4). In models adjusted for only one exposure variable and the potential confounders (Unadjusted Estimate) none of the exposure variables were significantly associated with pathogen diversity. In a model adjusted for all the exposure variables, location type (urban versus rural) and substrate type (soil versus water), Drainage Ditch (RR=1.39, 95%

CI 1.07, 1.81) and Uncollected Garbage (RR= 1.55, 95% CI 1.19, 2.03) were significantly associated with pathogen diversity.

4.4. Discussion

Urban and rural settings have distinct environmental risk profiles that can differ by contamination levels as well as by pathogen type. Pathogen-level contamination data is an essential element towards identifying important environmental transmission pathways. Water samples from our urban site had higher pathogen gene counts and a more diverse pathogen make up than samples from our rural sites (Figure 4-3 and Table 4-2). Within these urban areas, samples from drainage ditches (compared to stagnant water) and uncollected garbage within 50m of the sample collection point, were stronger environmental predictors of increased bacterial contamination (measured by pathogenic bacterial gene counts and pathogen diversity) than proximity to toilets or the presence of animals or animal feces (Tables 4-3 and 4-4). Elevation of the sample collection point was also a significant predicator of pathogenic bacterial gene counts. Viral gene counts, however, were not associated with any collection point characteristics (Table 4-3). Similarly, soil samples had a markedly different pathogen make up to water samples, with protozoa and viruses predominating in soil and bacteria in water (Figure 4-3). We also found that direct pathogen measurements provided more detailed information on environmental risk factors associated with pathogen contamination than our indicator organism.

Higher pathogen loads in urban versus rural sites are often attributed to increased human density resulting in increased rates of pathogen shedding into the environment. Urban areas are also characterized by a more dynamic environment compared to rural areas, with human activity constantly altering the environment. An example of such activities would be the building of drainage ditches to channel wastewater away from dwellings and the creation of informal

garbage dumps. These two site characteristics: drainage ditches and uncollected garbage were predominantly present only in urban areas and were associated pathogenic bacterial gene counts and pathogen diversity. Both drainage ditches and uncollected garbage are characteristics of sites where the environmental pathogen pool is constantly replenished, hence the association with less persistent bacterial pathogens (32). The higher density of people interacting with the environment and the presence of sites with high pathogen concentrations, characterized for example by drainage ditches and uncollected garbage, may also result in higher pathogen transmission rates in urban areas, potentially explaining why urban children are more prone to diarrhea despite better long-term outcomes, such as stunting and weight gain (33,34). Another potential consequence of elevated urban transmission rates coupled with a constantly changing environment maybe the selection for more virulent pathogens and the transfer of virulence and antibiotic resistance genes between enteric and environmental bacterial strains (34–36). In addition, uncollected garbage results in a complex matrix of both organic and inorganic substrates, enhancing the persistence or even potentially enabling replication bacterial enteropathogens like *E.coli* as was observed with EPEC in algal *Cladophora* mats (37).

Our finding that drainage ditches contained more pathogens than stagnant water suggests that these ditches are acting as accumulators of contamination within the urban site and subsequently spreading these pathogens throughout the community. This finding highlights the interconnected nature of drains in urban areas and the ability of drains in urban areas to move fecal contamination both within and between communities (5). One reason that stagnant water may have lower measured bacterial contamination levels is the rapid pathogen decay from solar radiation at the top of the water column and the settling out of pathogens that occurs when water is not flowing, suggesting that drainage ditches provide a more accurate snapshot of pathogens

shed by a community at a point in time compared to stagnant water. The lack of association of drainage ditches with viral gene counts, suggests that stagnant water may be more important for the transmission of viral pathogens, such as rotavirus. While the predominant route of transmission for rotavirus is host-to-host, the presence of an environmental reservoir such as stagnant water may serve to seed and sustain outbreaks (38). In addition, our finding that no environmental site characteristics were associated with viral gene loads is consistent with what is reported in literature, especially with regard to sanitation interventions. Sanitation interventions have been shown to be effective in preventing bacterial and protozoal infections, while having negligible effects on viral infection, indicating the environmental risk factors eliminated by sanitation interventions, such as dilapidated latrines, are not associated with viral transmission (39–41).

Elevation can also play a role in pathogen dissemination through a community given that contamination will flow along a drainage ditch from higher to lower elevations. Low lying areas at the confluence of multiple contamination streams are therefore likely to have the highest pathogens loads. Our finding that even minor elevation differences are associated pathogenic bacterial gene counts shows that in dense urban environments, there may be a non-trivial difference in disease risk along the course of a contaminated water conduit. Downstream locations along water conduits will have accumulated pathogens from a larger area of a community compared to upstream locations (42). In addition, in upstream locations, water conduits may serve to reduce disease risk by carrying away pathogens, while downstream, conduits serve to increase disease risk by carrying pathogens into a downstream community that may not have previously been present.

High protozoal and viral pathogen loads in soil (Figure 4-4) may be driven by the longer persistence of protozoa and viruses (32). Similar to what was found by Pickering *et al.* (2012), our findings point to soil being an underappreciated pathogen reservoir (43). Not only would the soil reservoir result in direct pathogen-host contacts, but it could also re-contaminate other environmental substrates especially after heavy rain events (16). Soil is often not the focus of sanitation interventions, therefore failing to account for it as a potential pathogen reservoir could reduce the effectiveness of interventions (43) . We also observed that uncollected garbage was present near 70% of the soil sites with highest pathogen counts (Figure 4-4), reinforcing the importance of soil as a pathogen reservoir. Given the lack of a clear mitigation strategy to decrease soil contamination, reinforcing hygiene practices (both food and hand hygiene) should be promoted.

Our findings point to a more a more nuanced relationship between fecal indicators such as *E. faecalis* and pathogens. Fecal indicators in our study highlight the much higher contamination levels in urban areas compared to rural areas and that drainage ditches and uncollected garbage are characteristic of highly contaminated sites. Samples collected from drainage ditches were significantly associated with *E.faecalis* gene counts (RR=3.02, 95% CI 1.17, 7.77), similar to what was observed with pathogenic bacterial gene counts, (RR= 2.31, 95% CI 1.43, 3.72). Similarly, samples collected close to garbage were significantly associated with both *E. faecalis* and pathogenic bacterial gene counts (RR=3.23, 95% CI 1.07, 9.72, RR=1.98 95% CI 1.13, 3.45 respectively). However, the lack of association between *E. faecalis* loads and site elevation may indicate that even though indicators may be accurate predicators of contamination in our study, they are unable to capture more subtle dynamics such the impact of elevation on pathogen movement within communities. The utility of fecal indicators might have

to be considered by the scale and context that are used for. In large scale tropical settings, indicator organisms are indicative of overall fecal contamination, similar to what we observed (15). Future studies should consider exploring if there is a scale effect in the correlation between fecal indicators and pathogens or pathogen types.

The presence of domestic animals was not associated with a more diverse pathogen community, unlike in previous studies (4). Furthermore, even though we found that only 2.4% of urban households reported owning livestock compared to 46% of rural households, animal ownership may not accurately reflect the presence of domestic animals, with animals liable to wander in from surrounding areas. We also did not sample home environments and do not know if this lack of association holds in domestic settings. Domestic animals may only be important in contributors to the pathogen pool in domestic settings in rural areas and not to the overall pathogen community pool in urban areas, where contributions from animals may be minor in relation to overall community shedding in densely populated urban areas. Further studies are necessary to accurately evaluate the impact that domestic animals have on environmental pathogen loads.

Characterizing environmental transmission pathways through direct measurement of pathogens, rather than relying on indirect measures such as the presence of or access to water treatment and sanitary infrastructure, will improve our ability to mitigate human risks to environmental exposures. Urban areas often score high on access to WASH infrastructure, for example an estimated 91% of the population in our urban study sites was reported to have access to improved drinking water sources and estimated 94% had access to improved sanitation (23). However, we found multiple pathogen reservoirs that present risks that would not be mitigated by improved water and sanitation access. Interestingly, rural populations are exposed to

different pathogen types, which may suggest different mitigation strategies. Enteropathogen transmission may also be driven by environmental characteristics that are not captured by WASH metrics; for example, a study in a peri-urban neighborhood in Yangon found that acute diarrhea was significantly associated with unsafe waste disposal (44). A more detailed spatiotemporal analysis of environmental sources of contamination, including human and animals, will provide an important transformative shift how we conceptualize effective intervention strategies.

Table 4-1: Descriptive characteristics stratified by urban and rural and by community. Toilet facilities were classified as functional if they were constructed of durable materials and had a door for privacy. Dilapidated toilet facilities were made of recycled materials and did not provide adequate privacy.

	URBAN			**RURAL**			
	SLB, *N=32*	**Ward 7, *N=31***	**Yeokkan, *N= 32***	**Chaung Nyiko, *N=12***	**Kyahaonne, N=*10***	**Urban, *N=95***	**Rural, *N=22***
Altitude, m (SD)	7.73 (3.01)	6.00 (3.08)	7.44 (5.02)	5.58 (2.39)	4.36 (1.69)	7.06 (3.87)	5.00 (2.13)
Human/Animal Feces Present, n (%)	1 (3.12)	8 (24.24)	9 (26.47)	3 (25.00)	0.00 (0.00)	18 (18.18)	3 (13.04)
Domestic Animals Present, n (%)	13 (40.62)	26 (78.79)	18 (52.94)	12 (100.00)	10 (90.91)	57 (57.58)	22 (95.65)
Toilet Visible, n (%)	3.00 (9.38)	4.00 (12.12)	11.00 (32.35)	4.00 (33.33)	8 .00 (72.73)	18.00 (18.18)	12.00 (52.17)
Functional toilets, n (%)	3.00 (100.00)	3.00 (100.00)	4.00 (90.91)	4.00 (100.00)	3.00 (37.50)	17.00 (94.44)	7.00 (58.33)
Dilapidated toilets, n (%)	0.00 (0.00)	0.00 (0.00)	1.00 (9.09)	0.00 (0.00)	5.00 (62.50)	1.00 (5.56)	5.00 (41.67)
Children Visible, n (%)	26.00 (81.25)	26.00 (78.79)	30.00 (88.24)	4.00 (33.33)	9.00 (81.82)	82.00 (82.83)	13.00 (56.52)
Uncollected Garbage n (%)	2.00 (6.25)	15.00 (45.45)	4.00 (11.76)	0.00 (0.00)	0.00 (0.00)	21.00 (21.21)	0.00 (0.00)

Table 4-2: Comparison of geometric mean pathogen gene counts between urban and rural sampling locations and by substrate. Pathogens with statistically significant differences in gene counts by site are bolded.

WATER			
	Urban Mean (95% CI)	**Rural Mean (95% CI)**	**p-value**
E.faecalis Count (gene copies/ml)	**2.32E+04 (1.51E+04, 3.56+04)**	**4.33E+02 (2.37E+02, 7.92E+02)**	**8.73E-10**
Bacteria (gene copies/ml)	**4.31E+02 (3.16E+02, 5.88E+02)**	**1.41E+02 (1.39E+02, 1.42E+02)**	**2.06E-07**
Protozoa (gene copies/ml)	**3.50E+01 (2.87E+01, 4.25E+01)**	**2.60E+01 (2.60E+01, 2.60E+01)***	**1.02E-02**
Helminths (gene copies/ml)	1.90E+01 (1.80E+01, 2.00E+01)	1.93E+01 (1.69E+01, 2.21E+01)	5.40E-01
Viruses (gene copies/ml)	**1.40E+02 (9.50E+01, 2.03E+02)**	**1.63E+01 (1.08E+01, 2.45E+01)**	**3.39E-07**
Number of Pathogens per Sample	**5.42E+00 (4.89E+00, 6.00E+00)**	**3.01E+00 (2.76E+00, 3.28E+00)**	**3.98E-08**
SOIL			
	Urban Mean (95% CI)	**Rural Mean (95% CI)**	**p-value**
E.faecalis Count (gene copies/g)	**1.09E+05 (4.14E+04, 8.77E+04)**	**1.97E+02 (1.00E-0.2, 3,83E+06)**	**7.37E-03**
Bacteria (gene copies/g)	1.96E+02 (1.33E+02, 2.90E+02)	1.40E+02 (1.40E+02, 1.40E+02)*	6.43E-01
Protozoa (gene copies/g)	8.54E+01 (4.15E+01, 1.72E+02)	1.06E+02 (2.50E-01, 4.49E+04)	9.46E-01
Helminths (gene copies/g)	2.86E+01 (1.49E+01, 5.49E+01)	1.80E+01 (1.80E+01,1.80E+01)*	6.43E-01
Viruses (gene copies/g)	1.20E+02 (4.28E+01, 3.35E+02)	1.67E+02 (6.00E-02, 4.52E+05)	7.57E-01
Number of Pathogens per Sample	2.88E+00 (2.61E+00, 3.18E+00)	2.88E+00 (1.21E+00, 6.86E+00)	9.77E-01

Table 4-3: Associations between sample gene counts and collection point characteristics. In addition to collection point characteristics, all models were adjusted for location type (urban vs rural), sample type (soil vs water). Significant associations are bolded.

	Log10 Enterococci Gene Count			Log 10 Pathogenic Gene Count			Log 10 Bacterial Gene Count			Log 10 Viral Gene Count		
Risk Factor	**RR**	**95% CI**		**RR**	**95% CI**		**RR**	**95% CI**		**RR**	**95% CI**	
Elevation	1.03	0.92	1.15	1.02	0.95	1.09	**0.94**	**0.88**	**0.99**	1.12	0.98	1.24
Drainage Ditch												
No		REF			REF			REF			REF	
Yes	**3.02**	**1.17**	**7.77**	1.56	0.83	2.94	**2.31**	**1.43**	**3.72**	1.14	0.44	2.93
Presence of Animals												
No					REF			REF			REF	
Yes	0.76	0.33	1.77	0.85	0.49	1.50	0.76	0.49	1.16	0.81	0.35	1.90
Nearby Toilet												
No		REF			REF			REF			REF	
Yes	2.44	0.96	6.22	1.83	0.97	3.44	1.13	0.70	1.83	1.96	0.76	5.01
Presence of Animal Feces												
No		REF			REF			REF			REF	
Yes	1.08	0.39	2.99	1.49	0.75	2.94	0.70	0.41	1.20	1.88	0.68	5.23
Uncollected Garbage												
No		REF			REF			REF			REF	
Yes	**3.23**	**1.07**	**9.72**	1.39	0.67	2.90	**1.98**	**1.13**	**3.45**	0.99	0.33	2.98

Table 4-4: Association between sample pathogen diversity and collection point characteristics. All models were adjusted for location type (urban vs rural) and sample type (soil vs water) in addition to exposure variables. Significant associations are bolded.

	Unadjusted Estimate			Adjusted Estimate		
Risk Factor	**Risk Ratio**	**95% CI**		**Risk Ratio**	**95% CI**	
Elevation	1.00	0.96	1.04	1.00	0.96	1.04
Drainage Ditch						
No	Ref			Ref		
Yes	1.23	0.93	1.62	**1.39**	**1.07**	**1.81**
Presence of Animals						
No	Ref			Ref		
Yes	0.87	0.68	1.11	0.80	0.64	1.01
Toilet Visible						
No	Ref			Ref		
Yes	0.96	0.66	1.40	0.99	0.71	1.38
Presence of Animal Feces						
No	Ref			Ref		
Yes	1.02	0.79	1.32	0.88	0.67	1.16
Uncollected Garbage						
No	Ref			Ref		
Yes	1.22	0.93	1.59	**1.55**	**1.19**	**2.03**

References

1. Null C, Stewart CP, Pickering AJ, Dentz HN, Arnold BF, Arnold CD, et al. Effects of water quality , sanitation , handwashing , and nutritional interventions on diarrhoea and child growth in rural Kenya : a cluster-randomised controlled trial. Lancet Glob Heal . 2018;6(3):E316–29.
2. Luby SP, Rahman M, Arnold BF, Unicomb L, Ashraf S, Winch PJ, et al. Effects of water quality , sanitation , handwashing , and nutritional interventions on diarrhoea and child growth in rural Bangladesh : a cluster randomised controlled trial. Lancet Glob Heal . 2018;6(3):E302–15.
3. Humphrey JH, Mbuya MNN, Ntozini R, Moulton LH, Stoltzfus RJ, Tavengwa N V, et al. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. Lancet Glob Heal. 2019;7(1):e132–47.
4. Baker KK, Senesac R, Sewell D, Sen Gupta A, Cumming O, Mumma J. Fecal fingerprints of enteric pathogen contamination in public environments of Kisumu, Kenya, associated with human sanitation conditions and domestic animals. Environ Sci Technol. 2018;52(18):10263–74.
5. Berendes DM, de Mondesert L, Kirby AE, Yakubu H, Adomako, Lady, Michiel J, et al. Variation in E. coli concentrations in open drains across neighborhoods in Accra, Ghana: The influence of onsite sanitation coverage and interconnectedness of urban environments. Int J Hyg Environ Health. 2020;224(113433).
6. Government of Myanmar M of H. Study on Cause of Under-Five Mortality. 2014.
7. Government of Myanmar M, United States Agency for International Development U. Myanmar Demographic Household Survey 2015-16. 2017.
8. Theingi Win Myat, Hlaing Myat Thu, Ye Myint Kyaw, Khin Mar Aye, Mo Mo Win, Htin Lin, et al. Sentinel surveillance for rotavirus in children <5 years of age admitted for diarrheal illness to Yangon Children's Hospital, Myanmar, 2009–2014. Vaccine . 2018;36(51):7832–5.
9. Moe K, Hummelman EG, Oo WM, Lwin T, Htwe TT. Hospital-based surveillance for rotavirus diarrhea in children in Yangon, Myanmar. J Infect Dis. 2005;192(SUPPL. 1):111–3.
10. Japan International Cooperation Agency (JICA), Yangon City Development Committee (YCDC). A Strategic Urban Development Plan of Greater Yangon Planning Area. 2013.
11. Myint ST, Myint T, Aung W, Wai K. Prevalence of household drinking-water contamination and of acute diarrhoeal illness in a periurban community in Myanmar. WHO South-East Asia J Public Heal. 2015;4(1):62.
12. Tofiloski S. Geospatial Analysis of Water-Associated Infectious Diseases: Case of Myanmar. Central European University; 2017.
13. Ferguson AS, Layton AC, Mailloux BJ, Culligan PJ, Daniel E, Smartt AE, et al. Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater. Sci Total Environ. 2012;(431):314–22.
14. Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, et al. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Appl Environ Microbiol. 2005;71(6):3163–70.
15. Lamparelli CC, Pogreba-Brown K, Verhougstraete M, Sato MIZ, de Castro Bruni A, Wade TJ, et al. Are fecal indicator bacteria appropriate measures of recreational water risks in the tropics: A cohort study of beach goers in Brazil? Water Res. 2015;87:59–68.
16. Carlton EJ, Eisenberg JNS, Goldstick J, Cevallos W, Trostle J, Levy K. Heavy rainfall events and diarrhea incidence: The role of social and environmental factors. Am J Epidemiol. 2014;179(3):344–52.

17. Boithias L, Choisy M, Souliyaseng N, Jourdren M, Quet F, Buisson Y, et al. Hydrological Regime and Water Shortage as Drivers of the Seasonal Incidence of Diarrheal Diseases in a Tropical Montane Environment. PLoS Negl Trop Dis. 2016;10(12).
18. Few R, Lake I, Hunter PR, Tran PG. Seasonality, disease and behavior: Using multiple methods to explore socio-environmental health risks in the Mekong Delta. Soc Sci Med. 2013;80:1–9.
19. Kostyla C, Bain R, Cronk R, Bartram J. Seasonal variation of fecal contamination in drinking water sources in developing countries: A systematic review. Sci Total Environ . 2015;514:333–43.
20. Knee J, Sumner T, Adriano Z, Berendes D, de Bruijn E, Schmidt WP, et al. Risk factors for childhood enteric infection in urban Maputo, Mozambique: A cross-sectional study. PLoS Negl Trop Dis. 2018;12(11):1–19.
21. United Nations U. World Urbanization Prospects . Vol. 12. 2018.
22. Government of Myanmar M of I and P. The 2014 Myanmar Population and Housing Census Tanintharyi Region. Vol. 3. 2015.
23. Government of Myanmar M of LI and P. The 2014 Myanmar Population and Housing Census: Hlintheya Township Report. 2017.
24. Government of Myanmar M of LI and P. The 2014 Myanmar Population and Housing Census: Htantabin Township Report. 2017.
25. Peel MC, Finalayson BL, McMahon TA. Long-term rates of mass wasting in Mesters Vig, northeast Greenland: Notes on a re-survey. Hydrol Earth Syst Sci. 2007;11(3):1633–44.
26. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, et al. Evaluation of a droplet digital polymerase chain reaction format for dna copy number quantification. Anal Chem. 2012;84:1003–11.
27. Cao Y, Raith MR, Griffith JF. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. Water Res. 2015;70:337–49.
28. Nshimyimana JP, Martin SL, Flood M, Verhougstraete MP, Hyndman DW, Rose JB. Regional variations of bovine and porcine fecal pollution as a function of landscape, nutrient, and hydrological factors. J Environ Qual. 2018;47(5):1024–32.
29. Hites RA. Correcting for censored environmental measurements. Environ Sci Technol. 2019;53(19):11059–60.
30. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1).
31. Zeileis A. Object-oriented computation of sandwich estimators. J Stat Softw. 2006;16(9):1–11.
32. Eisenberg JNS, Cevallos W, Ponce K, Levy K, Bates SJ, Scott JC, et al. Environmental change and infectious disease: How new roads affect the transmission of diarrheal pathogens in rural Ecuador. Proc Natl Acad Sci USA. 2006;103(51):19460–5.
33. Fink G, Günther I, Hill K. Slum residence and child health in developing countries. Demography . 2018;51(4):1175–97.
34. Smith SM, Montero L, Paez M, Ortega E, Hall E, Bohnert K, et al. Locals get travellers' diarrhoea too: risk factors for diarrhoeal illness and pathogenic Escherichia coli infection across an urban-rural gradient in Ecuador. Trop Med Int Heal. 2019;24(2):205–19.
35. Boots M, Mealor M, Science S, Series N, Mar N. Local interactions select for lower pathogen infectivity. Science (80-). 2007;315(5816):1284–6.
36. Bull JJ. Virulence. Int J Org Evol. 1948;161(4083):162–3.

37. Jang J, Hur HG, Sadowsky MJ, Byappanahalli MN, Yan T, Ishii S. Environmental Escherichia coli: ecology and public health implications—a review. J Appl Microbiol. 2017;123(3):570–81.
38. Kraay ANM, Brouwer AF, Lin N, Collender PA, Remais J V., Eisenberg JNS. Modeling environmentally mediated rotavirus transmission: The role of temperature and hydrologic factors. Proc Natl Acad Sci. 2018;115(12):E2782–90.
39. Berendes D, Leon J, Kirby A, Clennon J, Raj S, Yakubu H, et al. Household sanitation is associated with lower risk of bacterial and protozoal enteric infections, but not viral infections and diarrhoea, in a cohort study in a low-income urban neighbourhood in Vellore, India. Trop Med Int Heal. 2017;22(9):1119–29.
40. Berendes D, Kirby A, Clennon JA, Raj S, Yakubu H, Leon J, et al. The influence of household- and community-level sanitation and fecal sludge management on urban fecal contamination in households and drains and enteric infection in children. Am J Trop Med Hyg. 2017;96(6):1404–14.
41. Odagiri M, Schriewer A, Daniels ME, Wuertz S, Smith WA, Clasen T, et al. Human fecal and pathogen exposure pathways in rural Indian villages and the effect of increased latrine coverage. Water Res. 2016;100:232–44.
42. Reis RB, Ribeiro GS, Felzemburgh RDM, Santana FS, Mohr S, Melendez AXTO, et al. Impact of environment and social gradient on Leptospira infection in urban slums. PLoS Negl Trop Dis. 2008;2(4):11–8.
43. 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. 2012;46:5736–43.
44. Myint S, KhinThet W, Khaymar M, Aye PS. Estimation of acute diarrhea and acute respiratory infections among children under five years who lived in a peri-urban environment of Myanmar. OSIR J . 2013;6(4):13–8.

CHAPTER V

Conclusion

Despite substantial progress in reducing global diarrheal disease mortality, multiple challenges remain in understanding how to effectively implement WASH interventions to reduce the burden of enteric infections. One such challenge is how to accurately and objectively evaluate the impact that interventions have at both the individual and population levels(1). Traditional assessments of sanitation effectiveness have relied on visually checking for the presence of a technology and on self-reported diarrhea. However, as the work presented in this dissertation shows, the advent of new molecular tools opens new avenues for assessing the impact of interventions, both in terms of their impact on environmental pathogen loads and on physiological and immunological responses. The work in this dissertation highlighted the following : 1) the use of biomarkers to measure physiological and immunological response to enteropathogen carriage; 2) the use of bacteriological analysis, high throughput sequencing, community surveys, and pathogen identification to understand the impact of utility water quality on vulnerable populations; and 3) the joint use of molecular methods to understand the association between environmental factors and pathogen loads and types in informal settlements.

A major motivation of my work has been to understand the impact that pathogens and interventions have at the individual level. This is the motivation behind Chapter 2 of this dissertation. There is a dearth of studies that bridge our theoretical understanding of pathogenesis and actual pathogen carriage in vulnerable populations, where co-infections are often the norm. Understanding the physiological and immunological impact of pathogen carriage can help not

only inform the design of interventions, but also the metrics that are needed to evaluate the effectiveness of interventions. We are starting to recognize that long-term, often asymptomatic carriage of multiple pathogens may be the norm in many vulnerable populations, but the physiological and immunological impacts of this carriage remain poorly understood. To date our understanding of long-term pathogen carriage has largely revolved around the idea of EED, which, as we are starting to understand, is likely a diffuse condition with multiple exposures and physiological responses resulting in an endpoint that falls under the umbrella term of EED(1–5). Chapter 2 of this dissertation aims to add to our understanding of how to develop biomarker panels that can be used to measure specific cellular and immunological changes that occur with pathogen carriage. The end goal of this work is the development of a panel of biomarkers that would not only serve as a metric for the effectiveness of WASH interventions, but also provide readouts on long term outcomes such as child growth(6,7). The goal of interventions is to reduce pathogen exposure and biomarker panels may provide more accurate measures on the effectiveness of interventions and potentially provide further insights into which transmission pathways were most impacted by the intervention(6). Though still logistically and financially challenging, such work may be imperative in the evaluation of future interventions.

While Chapter 2 of this dissertation was focused on the individual, Chapter 3 describes work done at the city level and integrated multiple studies to understand the impact of utility water on infant health in informal settlements. This work highlights the challenges faced by vulnerable populations in urban areas(4,8,17,18,9–16). The impact of WASH access is typically evaluated using metrics developed for rural areas, where the bulk of studies have been conducted. However, as this dissertation shows, the shortcomings faced by urban dwellers are systemic and when effectiveness is evaluated at the individual level, results are hard to

contextualize. This is especially true given that urban populations are likely to score high on metrics such as ease of water access and the use of premise plumbing(19–22). An understanding of urban WASH requires a holistic evaluation, designed to capture both systemic failures at the city level and individual behaviors and habits. The challenge in urban areas may therefore be ensuring that existing utility systems keep up with population pressure and meet required levels of chemical and bacteriological quality. In addition, urban areas are highly contaminated, and eliminating all pathogen transmission pathways without transformative changes, may be near impossible(13). This necessitates that urban WASH adhere to a multi barrier approach (MBA) and place as many barriers as possible between the individual and the untreated source water taken up by urban utilities. The MBA would also a require an in-depth evaluation of the effectiveness and feasibility of different in-home water treatment regimes. As this dissertation demonstrated, pathogen contamination and water quality may render some commonly used in-home water treatment methods null.

Chapter 4 of this dissertation illustrates how new molecular methods are helping us better understand the nuances of associations between the environment and pathogen loads. This enables us to better quantify the risks associated with different environmental exposures. The work in Chapter 4 evaluated environmental characteristics associated with pathogen loads in informal settlements. We show that water can disseminate bacterial pathogens and how uncollected garbage results in localized points with high bacterial loads. In addition, our work also reinforces that environmental characteristics associated with high bacterial loads are not associated with high viral loads. Viruses are more persistent in the environment, resulting in viral loads decoupling from bacterial loads. The implication of this is that interventions that eliminate bacterial pathogens may not necessarily be effective in eliminating viral transmission. The fact

that interventions may not be effective in eliminating the transmission of all pathogens, also reinforces the need for more accurate measures, such as pathogen assessments, in the evaluation of intervention effectiveness(1). In addition, given the differences in urban and rural environments, understanding how the environment impacts pathogen loads in both rural and urban areas will be key in developing tailored interventions.

This dissertation couples' state-of-the-art molecular techniques with traditional epidemiological work. To continue to build on our past successes as a field, molecular methods within a systems-framework will be imperative. Continual reliance on traditional assessments will only add to unexpected results that will be hard to put into context. As stated at the beginning of this dissertation, diarrheal disease is a consequence of poverty and systemic neglect, and as a global community we largely know what measures need to be taken to solve these. The barriers to implementing such measures are unfortunately all too often financial and political, and not technological. However, as we head into the mid-point of the 21st century we face new challenges such as climate change and, in many places, collapsing ecosystems(23). To adapt to these challenges, public health and epidemiology will have to continually add to its suite of tools. With the advent of multiple 'omics' technologies, the challenge may not be a lack of tools, but rather how to adapt these new tools and technologies to epidemiological studies and how to use the data from them to inform interventions and mitigation strategies.

Finally, I hope that the work presented in this dissertation provides pathways to addressing some of the inequities in global health. Global health practice is littered with well-meaning, but ultimately ineffectual interventions. Among the litany of reasons for these failures is that, all too often, interventions are a solution looking for a problem and do not intervene at the most effective point. The work in this dissertation shows that we now have the technology to

accurately measure multiple disease transmission pathways and gauge which pathway to intervene on. Accurately assessing the transmission potential of pathways would enable us to not only design interventions, but also accurately gauge their effectiveness, both in the short and long term. Hopefully, this dissertation contributes to the knowledge base as we seek to build a more equitable world where access to high quality healthcare, clean water, and sanitation are basic human rights.

References

1. Goddard FGB, Ban R, Barr DB, Brown J, Cannon J, Colford JM, et al. Measuring environmental exposure to enteric pathogens in low-income settings: review and recommendations of an interdisciplinary working group. Environ Sci Technol. 2020;54(19):11673–91.
2. Fahim SM, Das S, Gazi MA, Mahfuz M, Ahmed T. Association of intestinal pathogens with faecal markers of environmental enteric dysfunction among slum-dwelling children in the first 2 years of life in Bangladesh. Trop Med Int Heal. 2018;23(11):1242–50.
3. Menzies IS, Zuckerman MJ, Nukajam WS, Somasundaram SG, Murphy B, Jenkins AP, et al. Geography of intestinal permeability and absorption. 1999;483–9.
4. Ngure FM, Reid BM, Humphrey JH, Mbuya MN, Pelto G, Stoltzfus RJ. Water , sanitation , and hygiene (WASH), environmental enteropathy , nutrition , and early child development : making the links. Ann N Y Acad Sci. 2014;1308:118–28.
5. Crane RJ, Jones KDJ, Berkley JA. Environmental enteric dysfunction : An overview. Food Nutr Bull. 2015;36(1):76–87.
6. Arndt MB, Cantera JL, Id LDM, Kalnoky M, White N, Bizilj G, et al. Validation of the Micronutrient and Environmental Enteric Dysfunction Assessment Tool and evaluation of biomarker risk factors for growth faltering and vaccine failure in young Malian children. PLoS Negl Trop Dis [Internet]. 2020;1–24.
7. Arndt MB, Walson JL. Enteric infection and dysfunction—A new target for PLOS Neglected Tropical Diseases. PLoS Negl Trop Dis. 2018;12(12):1–6.
8. Yuan T, Vadde KK, Tonkin JD, Wang J, Lu J, Zhang Z, et al. Urbanization impacts the physicochemical characteristics and abundance of fecal markers and bacterial pathogens in surface water. Int J Environ Res Public Health. 2019;16(10):1–19.
9. Neiderud CJ. How urbanization affects the epidemiology of emerging infectious diseases. African J Disabil. 2015;5(1):1–9.
10. Pörtner CC, Su Y hsuan. differences in child health across rural, urban, and slum areas: evidence from India. Demography. 2018;55(1):223–47.
11. Tanner M, Harpham T, Tanner M. Urbanization and Health in Developing Countries — a Review of Some Trends. Urban Heal Dev Series. 2018;1–8.
12. Fink G, Günther I, Hill K. Slum Residence and Child Health in Developing Countries. Demography. 2018;51(4):1175–97.
13. Holcomb DA, Knee J, Sumner T, Adriano Z, de Bruijn E, Nalá R, et al. Human fecal contamination of water, soil, and surfaces in households sharing poor-quality sanitation facilities in Maputo, Mozambique. Int J Hyg Environ Health. 2020;226.
14. Aliu IR, Akoteyon IS, Soladoye O. Living on the margins: Socio-spatial characterization of residential and water deprivations in Lagos informal settlements, Nigeria. Habitat Int. 2021;107:102293.
15. Amin N, Rahman M, Raj S, Ali S, Green J, Das S, et al. Quantitative assessment of fecal contamination in multiple environmental sample types in urban communities in Dhaka, Bangladesh using sanipath microbial approach. bioRxiv. 2019;1–21.
16. Ercumen A, Arnold BF, Kumpel E, Burt Z, Ray I, Nelson K, et al. Upgrading a piped water supply from intermittent to continuous delivery and association with waterborne illness: A matched cohort study in urban India. PLoS Med. 2015;12(10):1–24.

17. Nyambe S, Agestika L, Yamauchi T. The improved and the unimproved: Factors influencing sanitation and diarrhoea in a peri-urban settlement of Lusaka, Zambia. PLoS One . 2020;15(5):1–19.
18. Sinharoy SS, Pittluck R, Clasen T. Review of drivers and barriers of water and sanitation policies for urban informal settlements in low-income and middle-income countries. Util Policy. 2019;60(August):100957.
19. Lee EJ, Schwab KJ. Deficiencies in drinking water distribution systems in developing countries. J Water Health. 2005;3(2):109–27.
20. Kumpel E, Nelson KL. Comparing microbial water quality in an intermittent and continuous piped water supply. Water Res . 2013;47(14):5176–88.
21. Coelho ST, James S, Sunna N, Abu Jaish A, Chatiia J. Controlling water quality in intermittent supply systems. Water Sci Technol Water Supply. 2003;3(1–2):119–25.
22. Ercumen A, Gruber JS, Colford JM. Water distribution system deficiencies and gastrointestinal illness: A systematic review and meta-analysis. Environ Health Perspect. 2014;122(7):651–60.
23. Hooper DU, Adair EC, Cardinale BJ, Byrnes JEK, Hungate BA, Matulich KL, et al. driver of ecosystem change. Nature. 2012;486(7401):105–8.

APPENDIX

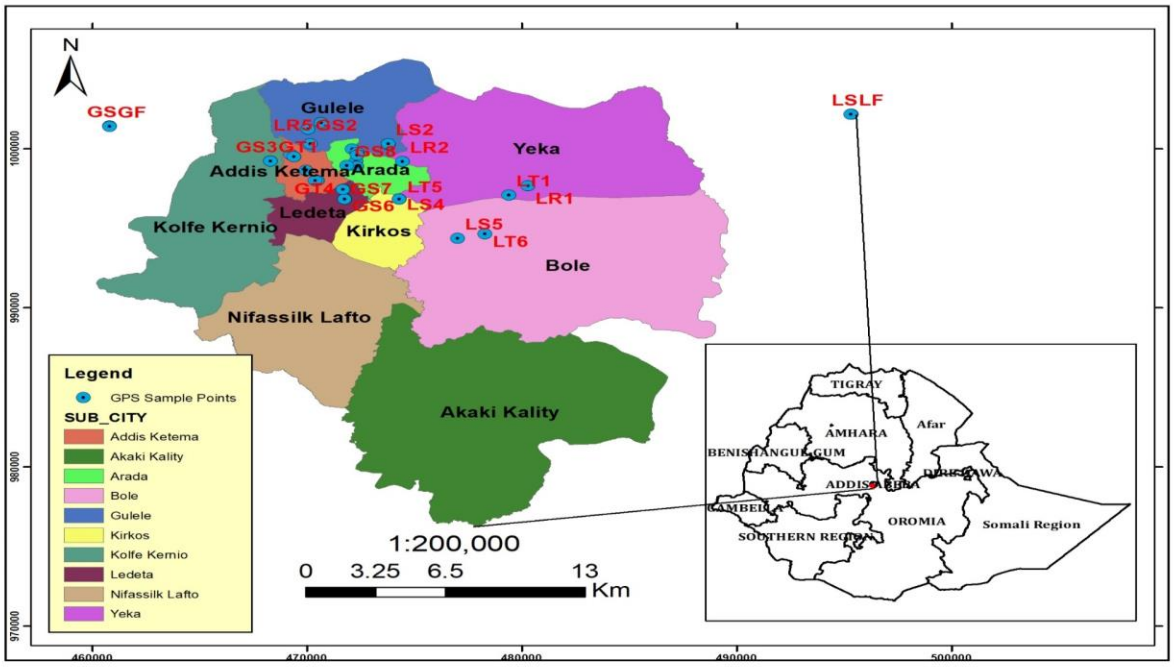


Figure A-1: Map of collection points for the bacteriological evaluation of utility supplied water in Addis Ababa, Ethiopia. G= Gefersa distribution system, L= Legedadi distribution system. Samples were collected at the following points: source taken up by the treatment plants (LS and GS, n=1 each); finished water before entering into distribution lines (LF and GF, n=1 each); large utility-maintained tanks (LR, n=7 and GR, n=2); taps inside buildings (LT, n=8 and GT, n=4); and storage tanks at buildings (LBS, n= 5 and GBS, n=8). Map was created and samples were collected by Bayable Atnafu Kassa, Addis Ababa University.

Table A-1: Household characteristics in study sites in the greater Yangon area, Myanmar. Data provided by Ther Aung, Carolina Population Center, University of North Carolina at Chapel Hill.

	Survey period (September-October 2019)						
	SLB (N=149)	**Ward 7 (N=150)**	**Yeokkan (N=151)**	**Urban (N=450)**	**Rural^ (N=150)**	**All (N=600)**	**Min-Max**
HOUSEHOLD RELATED							
Family size (person)	4.05 (1.59)	3.89 (1.89)	4.05 (1.80)	4.00 (1.76)	4.43** (1.79)	4.11 (1.78)	1-15
Age of household head (years)	43.48 (14.23)	47.90 (15.49)	43.02 (13.07)	44.80 (14.43)	51.49*** (12.86)	46.47 (14.34)	19-90
Female headed household (0/1)	0.201	0.307	0.205	0.238	0.160***	0.218	0-1
Head education							
None (0/1)	0.054	0.100	0.126	0.093	0.180***	0.115	0-1
Primary (0/1)	0.396	0.327	0.391	0.371	0.633***	0.437	0-1
Lower secondary (0/1)	0.315	0.267	0.265	0.282	0.107***	0.238	0-1
Upper secondary & above (0/1)	0.235	0.307	0.219	0.253	0.080***	0.210	0-1
Occupancy type							
Owner occupied (0/1)	0.315	0.407	0.338	0.353	0.980***	0.510	0-1
Tenancy (rent paid) (0/1)	0.658	0.580	0.411	0.549	0.007***	0.413	0-1
Tenancy (no rent paid) (0/1)	0.027	0.013	0.139	0.060	0.013**	0.048	0-1
Finished wall (0/1)	0.812	0.713	0.583	0.702	0.493***	0.650	0-1
Finished floor (0/1)	0.698	0.547	0.397	0.547	0.133***	0.443	0-1

	Survey period (September-October 2019)						
	SLB (N=149)	**Ward 7 (N=150)**	**Yeokkan (N=151)**	**Urban (N=450)**	**Rural^ (N=150)**	**All (N=600)**	**Min-Max**
Factories within 5 km of household (number)	0.90 (1.28)	0.61 (1.11)	2.79 (1.34)	1.44 (1.57)	0.06*** (0.33)	1.09 (1.50)	0-7
Informal (0/1)	0.295	0.120	0.523	0.313	0.000***	0.235	0-1
Livestock owned (0/1)	0.027	0.027	0.020	0.024	0.460***	0.133	0-1
Pig (number)	0.01 (0.08)	0.02 (0.25)	0.05 (0.46)	0.03 (0.31)	0.55*** (1.39)	0.16 (0.77)	0-12
Chicken (number)	0.08 (0.64)	0.40 (4.90)	0.01 (0.08)	0.16 (2.85)	3.24*** (7.79)	0.93 (4.79)	0-60
Duck (number)	0.00 (0.00)	0.20 (2.45)	0.00 (0.00)	0.07 (1.41)	7.98*** (34.69)	2.05 (17.68)	0-300
Cattle (number)	0.02 (0.25)	0.03 (0.29)	0.00 (0.00)	0.02 (0.22)	0.05 (0.44)	0.03 (0.29)	0-5
WATER & SANITATION							
Main source of drinking water							
Private/public piped water (0/1)	0.010	0.082	0.034	0.031	0.007	0.025	0-1
Bore hole/well (0/1)	0.000	0.000	0.017	0.009	0.048***	0.018	0-1
Rainwater (0/1)	0.010	0.010	0.017	0.020	0.408***	0.115	
Cart/tank delivery (0/1)	0.153	0.163	0.419	0.220	0.000***	0.166	0-1
Surface water (0/1)	0.010	0.000	0.009	0.011	0.517***	0.135	0-1
Bottled water (0/1)	0.816	0.735	0.496	0.705	0.014***	0.537	0-1

	Survey period (September-October 2019)						
	SLB (N=149)	**Ward 7 (N=150)**	**Yeokkan (N=151)**	**Urban (N=450)**	**Rural^ (N=150)**	**All (N=600)**	**Min-Max**
Main source of water not for drinking							
Private/public piped water (0/1)	0.327	0.224	0.248	0.286	0.184**	0.261	0-1
Bore hole/well (0/1)	0.663	0.633	0.513	0.585	0.673*	0.606	0-1
Rainwater (0/1)	0.000	0.020	0.009	0.007	0.007	0.007	0-1
Cart/tank delivery (0/1)	0.153	0.163	0.419	0.220	0.000***	0.166	0-1
Surface water (0/1)	0.010	0.000	0.009	0.011	0.517***	0.135	0-1
Toilet facility							
No facilities/bush (0/1)	0.000	0.000	0.017	0.004	0.000	0.003	0-1
Flush to septic tank (0/1)	0.490	0.410	0.237	0.389	0.233***	0.350	0-1
Flush to pit latrine (0/1)	0.388	0.495	0.508	0.471	0.553*	0.492	0-1
Flush somewhere else (0/1)	0.000	0.010	0.042	0.013	0.007	0.012	0-1
Pit latrine with slab (0/1)	0.082	0.086	0.085	0.084	0.093	0.086	0-1
Pit latrine without slab/open pit (0/1)	0.041	0.000	0.000	0.009	0.087***	0.028	0-1
Hanging toilet (0/1)	0.000	0.000	0.110	0.029	0.027	0.028	0-1

	Survey period (September-October 2019)						
	SLB (N=149)	**Ward 7 (N=150)**	**Yeokkan (N=151)**	**Urban (N=450)**	**Rural^ (N=150)**	**All (N=600)**	**Min-Max**
Toilet shared (0/1)	0.469	0.612	0.513	0.558	0.034***	0.430	0-1

^ statistical significance between rural and urban households. *p<0.05; **p<0.01; ***p<0.001