Distribution Agreement

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

Signature:

_____	_____
Frederick Goddard	Date

Enteric pathogen exposure and child health in low-income settings

By

Frederick G.B. Goddard

Doctor of Philosophy

Environmental Health Sciences

_____

Thomas F. Clasen, JD PhD
Advisor

_____ _____

Joe Brown, PhD
Committee Member

Howard H. Chang, PhD
Committee Member

_____ _____

Amy J. Pickering, PhD
Committee Member

Jeremy A. Sarnat, PhD
Committee Member

Accepted:

_____

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

_____

Date

Enteric pathogen exposure and child health in low-income settings

By

Frederick G.B. Goddard

B.Eng., University of Bristol, 2013

M.S., Stanford University, 2015

Advisor: Thomas F. Clasen, JD PhD

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Environmental Health Sciences.

2020

Abstract

Enteric pathogen exposure and child health in low-income settings

By Frederick G.B. Goddard

Background. Infections with enteric pathogens impair gastrointestinal function, causing diarrheal disease and nutritional deficits, the burden of which is highest among young children in low-income countries. Methods to estimate exposure to enteric pathogens suffer from a number of shortcomings, including the use of indicators of fecal contamination as proxies for enteric pathogens and that most exposure assessments estimate aggregate exposure (i.e. at the household- or community-level) instead of personal exposure. This dissertation sought to generate evidence to inform method development for improved measures of enteric exposure. **Methods.** The first aim estimated enteric exposure-health outcome relationships using household-level indicators of fecal contamination. We used an individual participant data (IPD) meta-analysis approach by requesting data from research identified through a systematic review. The second aim evaluated potential sources of measurement error in these proxy measures of exposure, by adapting an air pollution epidemiology exposure measurement error framework to drinking water quality. The third aim assessed the utility of saliva to estimate past exposure by measuring salivary antibody concentrations and comparing them to enteric pathogens detected in matched stool samples. **Findings.** Our IPD analyses found that fecal contamination in drinking water is associated with both diarrhea and impaired linear growth, and also implicated contaminated hands in diarrhea and contaminated fomites in stunted growth. The measurement error simulations suggested that household-level exposure assessments that do not consider exposure in the community may attenuate true water quality-diarrhea associations, particularly for older children. We also found that using single water quality measures, thereby not accounting for temporal variability in water quality, may attenuate the true effects of water quality on child growth. The saliva research found lower antibody concentrations in children experiencing higher numbers of concurrent enteric infections, and highlighted the need for further validation of salivary diagnostics for the detection of pathogen-specific antibodies. **Conclusions.** The findings from research conducted for this dissertation highlight the shortcomings of enteric exposure assessments. They also underline opportunities to improve methods used to estimate enteric exposure, which could subsequently contribute to informing policies and interventions designed to reduce the burden of enteric infections in underrepresented and underserved populations.

Enteric pathogen exposure and child health in low-income settings

By

Frederick G.B. Goddard

B.Eng., University of Bristol, 2013

M.S., Stanford University, 2015

Advisor: Thomas F. Clasen, JD PhD

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Environmental Health Sciences.
2020

Acknowledgements

I would like to express my gratitude to the many people who have provided support during my doctoral training, without whom this journey would not have been nearly as enjoyable or fulfilling. First, I would like to thank my family and friends for their support, in particular my parents, Evamarie Bange and Richard Goddard, and my partner Jenna Swarthout. My mother, or Mami as I like to call her, has always stood by my side, including through these past ten years of academic training and no matter how far I am from home. I would not have had the same confidence to go on this journey without my dad, who has always been open to and supportive of my decisions. Jenna drives me to be a better scientist and has spent countless hours providing feedback on my work, as well as my written and verbal communication. My friends have provided much needed distraction over the years, most notably by helping me discover my love for baking bread.

Second, let me acknowledge the chair of my dissertation committee, Prof. Thomas Clasen. I first met Tom when I interviewed at Emory. I have not experienced an interviewer that was as prepared and engaged as Tom was that day, which gave me confidence that, under his guidance, Emory was the right place for this training. I am grateful to Tom for finding the right balance in pushing me to do better science when I needed it and supporting my decisions on the rare occasions I was on the right track. Your ambition is unparalleled and I am sorry we did not invent the Goddard method for enteric exposure assessments as you had initially hoped. Thank you for having a unique capability to demonstrate such trust in your students and for empowering us to conduct independent research.

Third, I would like to highlight each of my committee members who have all played a distinct role in my development. Dr. Amy Pickering has been mentoring me for the past six years going back to my time as a Masters student and is one of the main reasons I developed a keen interest in research and chose to pursue this degree. Dr. Jeremy Sarnat helped make my decision to pursue this degree at Emory an easy one and always encouraged me to go outside of my comfort zone. Dr. Joe Brown provided me with my first taste of working in a laboratory and showed great patience in allowing me to complete one of my dissertation

chapters at the Georgia Institute of Technology. Dr. Howard Chang is one of the best teachers I have come across and has the rare ability to make statistics both understandable and exciting to non-statisticians like me.

Fourth, I cannot think of a more supportive environment for PhD students to thrive and be successful than the Environmental Health department at Emory. I have enormous gratitude to the department and the many friendly faces that made coming to work every day a fun and engaging experience. To my fellow Environmental Health Science PhD students, thank you for joining me in this journey and for heckling me whenever I started taking myself too seriously. I hope I have been able to make your doctoral experience at least partly as enjoyable as you have made mine.

Finally, I would like to acknowledge the Rose Salamone Gangarosa Endowed Chair in Sanitation and Safe Water, the Laney Graduate School and the Bill & Melinda Gates foundation for providing financial support for this work.

Table of Contents

List of Figures

List of Tables

1 Chapter 1: Introduction

1.1 Motivation

Rapid population growth and urbanization have led to an increase of fecal loading in the environment in low-income countries.[1] Conventional engineering solutions used in high-income countries to manage fecal sludge involve expensive and well-maintained infrastructure, including piped sewerage and centralized treatment. Resource constrained governments in low-income countries have often not been able meet the demand for this infrastructure, leading to inadequate sanitation coverage and substandard fecal sludge containment, transport and treatment that contribute to high fecal loading in the environment. Enteric infections associated with exposure to fecal contamination impair gastrointestinal function, causing diarrheal diseases and stunting due to poor nutrient absorption, the burden of which is highest in children under the age of five living in low-income settings.[2] enteric infections represent the third leading cause of death among children under five, accounting for approximately 589,000 deaths in 2017.[3] An estimated 151 million children were stunted in 2018, with 91% of those children living in low- or lower-middle-income countries.[4] An estimated 62% of diarrheal deaths and 16% of malnutrition among children under five are linked to enteric pathogen exposure from poor drinking water, sanitation and hygiene (WaSH) behaviors.[5]

For the purpose of this dissertation, enteric pathogens are broadly defined as microorganisms transmitted via the fecal-oral route that can cause gastrointestinal infections, leading to acute (i.e. diarrheal disease) and chronic infectious disease outcomes (i.e. persistent intestinal inflammation, growth faltering, impaired cognitive ability).[6] Enteric pathogens include bacterial, viral and protozoan pathogens, with fungi and helminths increasingly receiving more attention as neglected sources of tropical infectious disease.[7,8] The Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) estimated that the four leading aetiologies for diarrhea mortality in 2016 for children under the age of five were infections with Rotavirus, *Shigella*, Adenovirus and *Vibrio cholerae*.[9] The Global Enteric Multicenter Study (GEMS), a matched case-control study that enrolled children aged 0-59 months in the Gambia, Mali, Mozambique,

Kenya, India, Bangladesh and Pakistan, found that most attributable cases of moderate-to-severe diarrhea were due to infections with Rotavirus, *Cryptosporidium*, *Escherichia coli* producing heat-stable toxin (ST-ETEC) and *Shigella*.[10] The Etiology, Risk Factors, and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development Project (MAL-ED), a birth cohort study with 24 month follow-up in Bangladesh, India, Nepal, Pakistan, South Africa Tanzania, Brazil and Peru, found that Norovirus, Rotavirus, *Campylobacter* spp and Astrovirus were associated with the highest attributable fraction of diarrhea in the first two years of life.[11] These findings indicate that the disease severity from enteric pathogen infections depends in part on the specific pathogen, and evidence from GEMS and MAL-ED suggests that there is substantial variation in infections with specific pathogens by location and some variation in infections and attributable burdens of those infections by age.[10,11]

Exposure to enteric pathogens is conditional on a complex source-to-host fecal-oral transmission pathway. This includes: 1) the shedding of enteric pathogens in human or animal feces (the source); 2) the migration and transformation of the pathogen in the environment leading to a specific concentration in different environmental reservoirs; 3) the host interacting with those reservoirs; and 4) the pathogen entering the host (i.e. through oral ingestion). Whether ingestion of enteric pathogens then leads to adverse infectious disease outcomes depends on a number of factors, including the infectivity of the ingested pathogens and the ability of the host to mount an immune response. Figure 1-1 provides a summary enteric exposure schematic, using viral source-to-host transmission through the drinking water pathway as an example.

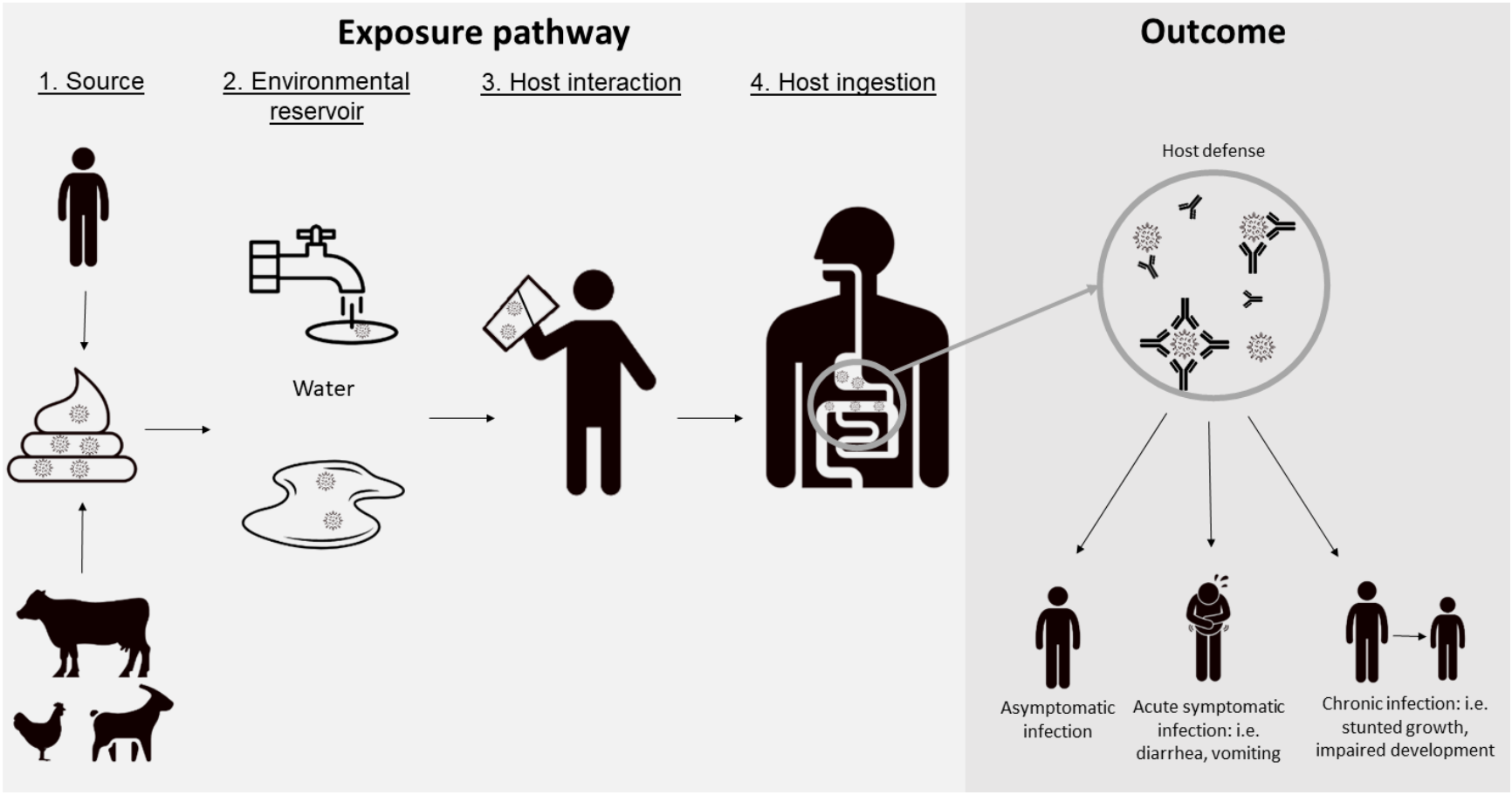


Figure 1-1: Enteric pathogen exposure – Example of viral exposure through the drinking water pathway

Methods to characterize human exposure to enteric pathogens have not been advanced in the same way as they have in other areas of environmental health, such as exposure to air pollution or chemical toxicants, perhaps in part because the resources are not available in the settings where enteric infections carry a disproportionally high burden of disease. Exposure science was born out of a need to characterize industrial and occupational exposure to air pollutants and chemical toxicants.[12] This need arose in high-income countries after the industrial revolution led to a combination of rapid urbanization and proliferation of factories in urban centers exploiting processes involving combustion to fuel their productions. Perhaps one of the most notable examples of this sudden increase of population exposure to high levels of air pollution is the London smog incident, which caused an estimated 12,000 casualties in 1952.[13] Thereafter, air pollution exposure scientists developed methods that advanced exposure assessments, ranging from stationary equipment and satellite imaging to measure ambient exposures to portable air pollution monitors that can provide personal exposure data, both cumulative and in real time.[14] In addition, these approaches can measure disease-causing agents, such as fine particulate matter and carbon monoxide, not indicators as proxies. Chemical exposure assessments have been advanced to include biomonitoring, i.e. using human biological matrices such as urine and serum with mass spectrometry-based methods, that have resulted in many population-based data reports on human chemical exposure in the United States, Canada and Korea.[15–17]

Exposure data can be an effective tool to inform policy to reduce population health burdens from environmental contaminants, i.e. what gets measured gets managed.[18] In the United States, for example, air quality data from ambient air-monitoring sites are used widely to establish regulatory limits on ambient air quality.[19] The European Union uses chemical exposure assessments to conduct risk assessments for chemical toxicants and inform the European regulation on Registration, Evaluation, Authorization, and restriction of Chemicals (REACH).[20] Method development for the monitoring of exposure to enteric pathogens has largely been focused on monitoring of recreational water quality and drinking water treatment processes for large, centralized drinking water supply systems.[21] Methods to

characterize enteric pathogen exposure from a variety of possible exposure pathways in the household- or community-environment has arguably not received the same resources and attention, likely in no small part because populations that are most commonly exposed at this level and most vulnerable to these exposures are young children in low-income countries.

Other reasons exposure assessments for enteric pathogens have not been advanced in the same way as they have for other contaminants, include the technological difficulties of detecting low numbers of diverse pathogens in different environmental compartments,[22] and the mixed evidence of an association between indicators of fecal contamination and specific enteric pathogens.[23,24] The pathogenicity, diversity and evolution differ by pathogen as well their ability to amplify and die-off both in the environment and inside the host.[25] Enteric exposure assessments need to consider multiple possible fecal-oral transmission pathways, with common pathways represented by the F-diagram[26] – fingers, flies, food, fluids (water sources) and fields (soil), with fomites an additional pathway often considered important.[27–29] The biological relevance of enteric exposures is also complicated by host susceptibility mediating dose-response relationships. Infection with a pathogen might alter future responses to that same infection due to acquired immunity,[30] and the ability to mount an immune response to enteric infections is influenced by gastrointestinal health, such as the diversity of the gut microbiome[31] and environmental enteric dysfunction (EED) from repeated enteric pathogen exposure.[32]

1.1.1 The need for improved exposure methods: the WaSH example

Systematic reviews of WaSH interventions, designed to demonstrate the health benefits of reducing enteric pathogen exposure, have generally found them to be protective against WaSH related diseases, including diarrhea,[33] soil-transmitted helminthiasis,[34] trachoma[35] and malnutrition,[36,37] however the evidence is mostly from observational studies. Recent experimental field evaluations of these interventions found either no evidence of health benefits,[38–40] a reduction in diarrhea but no improvement in child growth,[41] or improved growth but no impact on diarrhea.[42,43] These mixed results have focused greater attention on the need for more rigorous exposure assessment, in part to explain why effects of

WaSH interventions are realized in some trials and not others. A recent consensus statement among researchers in WaSH recommends that interventions need to radically reduce fecal contamination in the environment to achieve more consistent child health benefits,[44] though which microbes and pathways are most important – and the necessary reductions needed to achieve health impacts – may be highly context-specific. Filling these gaps requires greater attention on more rigorous exposure assessment.

The theory of change underlying the impact of WaSH interventions on enteric health outcomes is that an intervention will prevent disease if it a) is capable of reducing exposure to enteric pathogens, (b) is introduced into a vulnerable population, (c) achieves high levels of coverage and use, and (d) reduces population exposure to enteric pathogens. WaSH studies have traditionally measured some of the steps along this theory of change,[45] but only a few have actually attempted to assess the impact of an intervention on enteric exposure along the transmission pathways targeted by the interventions. Some of the null findings from recent WaSH interventions are consistent with the WaSH theory of change, reporting null effects from potentially effective interventions delivered to a vulnerable population when coverage and uptake were low.[38,40,46,47] However, other trials have reported null effects on diarrhea[39,48] and/or stunting[39,41,48] despite higher levels of coverage and use, while others have reported protective effects on stunting (but not diarrhea) with high levels of coverage and use, especially from reductions in open defecation.[42,43,49]

Results from exposure assessments included in a subset of these evaluations have raised important methodological questions about the utility of current methods for assessing child exposure to enteric pathogens (Figure 1-2). Evaluations led by Hartinger and Luby (sanitation arm) found positive intervention effects on child diarrhea, despite no reductions in measured household-level fecal contamination as a proxy for enteric pathogen exposure.[41,50] An evaluation led by Reese found positive intervention effects on child linear growth, despite no evidence of a reduction in fecal contamination in drinking water or on child hands.[51] A study led by Pickering also found positive intervention effects on child growth, despite no reduction in fecal contamination in drinking water; however, latrine fly presence

and observed human and animal feces did significantly decrease in the treatment group.[42] Evaluations led by Null (water and combined water, sanitation and hygiene interventions) and Luby (water intervention), found no improvements in child diarrhea or linear growth even though they found reductions in fecal contamination.[39,41] Evaluations led by Arnold, Clasen, Null (sanitation and hygiene arms) and Patil were consistent with the theory of change, in that they found no improvements in measured exposures and subsequently no improvements in primary health outcomes.[38–40,47] The hygiene and combined water, sanitation and hygiene arms in an evaluation led by Luby also included findings that were consistent with the theory of change, finding a reduction fecal contamination and improvements in health outcomes.[41]

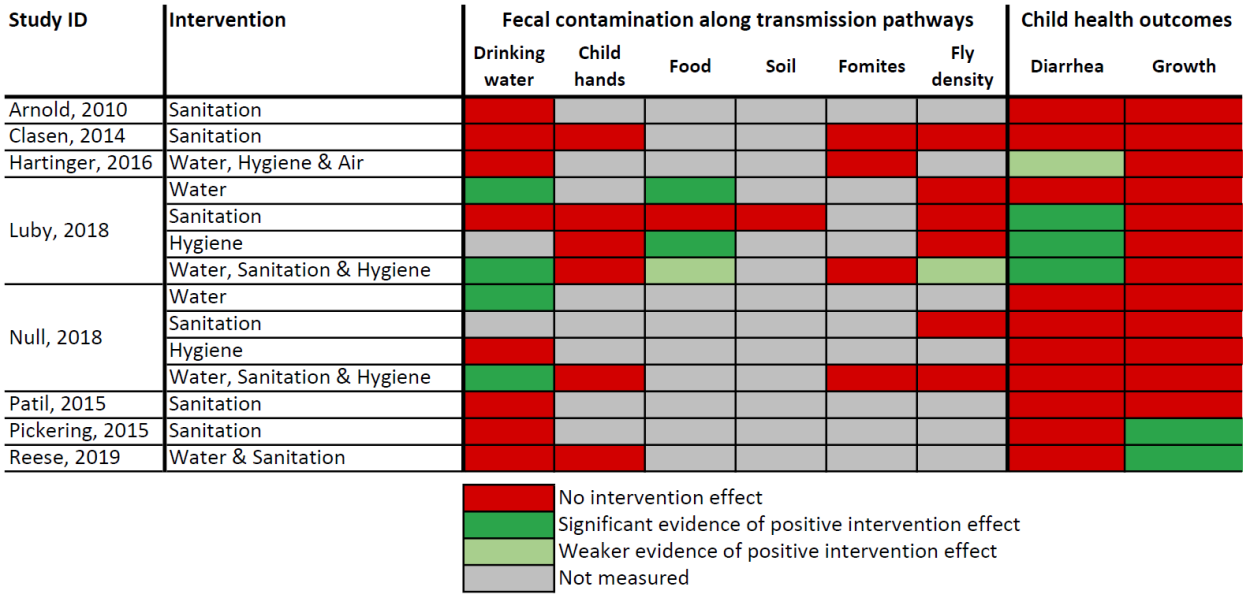


Figure 1-2: Summary of intervention effects from recent evaluations on fecal contamination along common transmission pathways (drinking water, child hands, food, soil, fomites and food preparation area fly density) and child diarrhea and stunting.

These exposure assessments measured fecal contamination along select transmission pathways using fecal indicators to estimate intervention effects on fecal contamination in the household and its surrounding environment. Whether these measures can serve as effective proxies for personal exposure to enteric pathogens depends on a number of assumptions. These assumptions include, but are not limited to, the (i) household-level exposure assessments are an effective proxy for individual-level exposure, (ii) fecal indicators are an effective proxy for enteric pathogens, (iii) key transmission pathways were captured in

the exposure assessment, and (iv) recovery of fecal indicators with the sampling and processing methods is reliable across different transmission pathways.

The inconsistencies between fecal contamination in the household environment and subsequent health outcomes are not seen as a challenge to the theoretical role of environmental mediation of fecal exposure on enteric infection and related sequellae.[52] However, the lack of a clear and consistent progression between exposure and health in several studies raises fundamental questions about current methods of exposure assessment, and especially the basic question about "how clean is clean enough" to realize reductions in infection and disease.

With this background, this dissertation sought to assess the relationships between current proxy measures of environmental enteric exposure and adverse child health outcomes, and make contributions to inform enteric exposure method development moving forward. There are a number of applications that could benefit from improved measures of enteric exposure. These include 1) identifying the dominant sources of enteric pathogen transmission to inform intervention design; 2) enabling potentially faster evaluation of interventions; and 3) applying new approaches to environmental surveillance of human infection prevalence and environmental threats.

1.2 Dissertation aims

This dissertation fulfills three core research aims. The first aim attempted to quantify exposure-outcome relationships using proxy measures of environmental enteric exposure, the second aim evaluated potential sources of error in those proxy measures, and the third aim assessed the use of salivary antibodies to estimate past exposure to enteric pathogens. This dissertation also included a supplemental aim in collaboration with a working group of interdisciplinary researchers to review methods used to estimate exposure to enteric pathogens and make recommendations on how to move enteric exposure method development forward. This document draws on that supplemental aim during the introductory and concluding chapters.

1.2.1 Research Aim 1

The first aim titled "*Fecal contamination of the environment and child health - A systematic review and meta-analysis using individual participant data*" sought to assess the relationship between household-level fecal contamination, as a proxy for individual-level enteric exposures, and adverse child health outcomes by addressing the following research questions:

1.1. Are levels of fecal contamination in the environment associated with diarrheal disease prevalence in children under the age of 5?

1.2. Are levels of fecal contamination in the environment associated with linear growth in children under the age of 5?

1.2.2 Research Aim 2

The second aim titled "*Exposure measurement error and the characterization of child exposure to fecal contamination in drinking water"* adapted an air pollution epidemiology exposure measurement error framework and evaluated potential sources of exposure measurement error from proxy measures of enteric exposure. The specific research questions for this aim were:

2.1. Is there evidence of exposure measurement error from in household-level exposure assessments that do not consider exposure outside of the household in the community?

2.2. Is there evidence of exposure measurement error from exposure assessments that estimate exposure at single time points compared to those with repeated longitudinal measures?

1.2.3 Research Aim 3

The third aim titled "*Child salivary SIgA and its relationship to enteric infections and EED biomarkers in Maputo, Mozambique*" sought to test the utility of saliva to both detect infections with specific pathogens as a measure of past exposure and more broadly as an alternative biological matrix to characterize gastrointestinal health. The research questions addressed by this aim were:

3.1. Are pathogen-specific acute salivary antibody responses associated with infection with those pathogens detected in matched stool samples?

3.2. What is the relationship between non-specific acute salivary antibodies and enteric infections as well as biomarkers of gut inflammation detected in matched stool samples?

1.3 References

1 Garenne M. Urbanisation and child health in resource poor settings with special reference to under-five mortality in Africa. Arch. Dis. Child. 2010; **95**: 464–8.

2 Kosek M, Bern C, Guerrant RL. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bull World Health Organ* 2003; **81**: 197–204.

3 Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017). 2018. http://ghdx.healthdata.org/record/global-burden-disease-study-2017-gbd-2017-reference-life-table (accessed Jan 27, 2020).

4 UNICEF, WHO WBG. Levels and trends in child malnutrition. 2018 https://www.who.int/nutgrowthdb/2018-jme-brochure.pdf (accessed June 13, 2019).

5 Prüss-Ustün A, Wolf J, Bartram J, *et al.* Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *Int J Hyg Environ Health* 2019; **222**: 765–77.

6 Kolling G, Wu M, Guerrant RL. Enteric pathogens through life stages. *Front Cell Infect Microbiol* 2012; **2**. DOI:10.3389/fcimb.2012.00114.

7 Bongomin F, Gago S, Oladele RO, Denning DW. Global and multi-national prevalence of fungal diseases—estimate precision. J. Fungi. 2017; **3**. DOI:10.3390/jof3040057.

8 Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ, Jacobson J. Helminth infections: The great neglected tropical diseases. J. Clin. Invest. 2008; **118**: 1311–21.

9 GBD. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis* 2018; **18**: 1211–28.

10 Kotloff KL, Nataro JP, Blackwelder WC, *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 2013; **382**: 209–22.

11 Platts-Mills JA, Babji S, Bodhidatta L, *et al.* Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet Glob Heal* 2015; **3**: e564–75.

12 Lioy PJ. Exposure science: a view of the past and milestones for the future. *Environ Health Perspect* 2010; **118**: 1081–90.

13 Bell ML, Davis DL, Fletcher T. A retrospective assessment of mortality from the London smog episode of 1952: the role of influenza and pollution. *Environ Health Perspect* 2004; **112**: 6–8.

14 Snyder EG, Watkins TH, Solomon PA, *et al.* The Changing Paradigm of Air Pollution Monitoring. *Environ Sci Technol* 2013; **47**: 11369–77.

15 CDC. National Report on Human Exposure to Environmental Chemicals. https://www.cdc.gov/exposurereport/ (accessed Dec 20, 2019).

16 Statistics Canada. The Canadian Health Measures Survey. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/environmental-contaminants/human-biomonitoring-environmental-chemicals/canadian-health-measures-survey.html (accessed Dec 20, 2019).

17 Kweon S, Kim Y, Jang M-J, *et al.* Data Resource Profile: The Korea National Health and Nutrition Examination Survey (KNHANES). *Int J Epidemiol* 2014; **43**: 69–77.

18 Castrucci BC, Rhoades EK, Leider JP, Hearne S. What gets measured gets done: An assessment of local data uses and needs in large urban health departments. *J Public Heal Manag Pract* 2015; **21**: S38–48.

19 US EPA. Reviewing National Ambient Air Quality Standards (NAAQS): Scientific and Technical Information. https://www.epa.gov/naaqs (accessed Aug 7, 2019).

20 Landberg HE, Hedmer M, Westberg H, Tinnerberg H. Evaluating the Risk Assessment Approach of the REACH Legislation: A Case Study. *Ann Work Expo Heal* 2019; **63**: 68–76.

21 The National Research Council. Indicators for Waterborne Pathogens. Washington, DC: National Academies Press (US), 2004 https://www.ncbi.nlm.nih.gov/books/NBK215658/ (accessed June 12, 2018).

22 Straub TM, Chandler DP. Towards a unified system for detecting waterborne pathogens. *J Microbiol Methods* 2003; **53**: 185–97.

23 Wu J, Long SC, Das D, Dorner SM. Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. *J Water Health* 2011; **9**: 265–78.

24 Hörman A, Rimhanen-finne R, Maunula L, Bonsdorff C Von, Torvela N. Indicator Organisms in Surface Water in and Indicator Organisms in Surface Water in Southwestern. *Appl Environ Microbiol* 2004; **70**: 87–95.

25 Gupta S, Maiden MCJ. Exploring the evolution of diversity in pathogen populations. *Trends Microbiol* 2001; **9**: 181–5.

26 Wagner EG, Lanoix JN. Excreta disposal for rural areas and small communities. Monogr. Ser. World Health Organ. 1958; **39**: 1–182.

27 Vujcic J, Ram PK, Hussain F, *et al.* Toys and toilets: Cross-sectional study using children's toys to evaluate environmental faecal contamination in rural Bangladeshi households with different sanitation facilities and practices. *Trop Med Int Heal* 2014; **19**: 528–36.

28 Torondel B, Gyekye-Aboagye Y, Routray P, Boisson S, Schimdt W, Clasen T. Laboratory development and field testing of sentinel toys to assess environmental faecal exposure of young children in rural India. *Trans R Soc Trop Med Hyg* 2015; **109**: 386–92.

29 Ngure FM, Humphrey JH, Mbuya MNN, *et al.* Formative Research on Hygiene Behaviors and Geophagy among Infants and Young Children and Implications of Exposure to Fecal Bacteria. *Am J Trop Med Hyg* 2013; **89**: 709–16.

30 Kotton CN, Hohmann EL. Enteric pathogens as vaccine vectors for foreign antigen delivery. *Infect Immun* 2004; **72**: 5535–47.

31 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**: ofx144.

32 Bourke CD, Berkley JA, Prendergast AJ. Immune Dysfunction as a Cause and Consequence of Malnutrition. Trends Immunol. 2016; **37**: 386–98.

33 Clasen TF, Alexander KT, Sinclair D, *et al.* Interventions to improve water quality for preventing diarrhoea. *Cochrane database Syst Rev* 2015; : Issue 10. Art No.: CD004794.

34 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**: e1001620.

35 Stocks ME, Ogden S, Haddad D, Addiss DG, McGuire C, Freeman MC. Effect of Water,

Sanitation, and Hygiene on the Prevention of Trachoma: A Systematic Review and Meta-Analysis. *PLoS Med* 2014; **11**: e1001605.

36 Dangour AD, Watson L, Cumming O, *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; : Issue 8. Art. No.: CD009382.

37 Freeman MC, Garn J V, Sclar GD, *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**: 928–49.

38 Clasen T, Boisson S, Routray P, *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* 2014; **2**: e645–53.

39 Null C, Stewart CP, Pickering AJ, *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**: 316–29.

40 Patil SR, Arnold BF, Salvatore AL, *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**. DOI:10.1371/journal.pmed.1001709.

41 Luby SP, Rahman M, Arnold BF, *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**: E302–15.

42 Pickering AJ, Djebbari H, Lopez C, Coulibaly M, Alzua ML. 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.

43 Reese H, Routray P, Torondel B, *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; **0**: 1–11.

44 Pickering AJ, Null C, Winch PJ, *et al.* The WASH Benefits and SHINE trials: interpretation of WASH intervention effects on linear growth and diarrhoea. *Lancet Glob Heal* 2019; **7**: e1139–46.

45 Turman-Bryant N, Clasen TF, Fankhauser K, Thomas EA. Measuring progress towards sanitation and hygiene targets: a critical review of monitoring methodologies and technologies. *Waterlines* 2018; **37**: 229–47.

46 Boisson S, Stevenson M, Shapiro L, *et al.* Effect of Household-Based Drinking Water Chlorination on Diarrhoea among Children under Five in Orissa, India: A Double-Blind Randomised Placebo-Controlled Trial. *PLoS Med* 2013; **10**: e1001497.

47 Arnold BF, Khush RS, Ramaswamy P, *et al.* Causal inference methods to study nonrandomized, preexisting development interventions. *PNAS* 2010; **107**: 22605–10.

48 Humphrey JH. Sanitation Hygiene Infant Nutrition Efficacy (SHINE) Trial in Zimbabwe. In: World BanK Webinar. 2018.

49 Spears D, Ghosh A, Cumming O. Open Defecation and Childhood Stunting in India: An Ecological Analysis of New Data from 112 Districts. *PLoS One* 2013; **8**: e73784.

50 Hartinger S, Lanata C, Hattendorf J, *et al.* Improving household air, drinking water and hygiene in rural Peru: a community-randomized–controlled trial of an integrated environmental home-based intervention package to improve child health. *Int J Epidemiol* 2016; **45**: dyw242.

51 Reese HE, Brown J, Pickering AJ, Mishra SR, Chang HH, Clasen TF. The effect of a combined household piped-water and sanitation intervention on fecal contamination of source water, drinking water, and children's hands: a matched cohort study in Odisha, India. .

52 Cumming O, Arnold BF, Ban R, *et al*. The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: a consensus statement. *BMC Med* 2019; **17**: 173.

2 **Chapter 2:** Estimating enteric exposure: A review of approaches

Manuscript authors: Frederick G.B. Goddard, Radu Ban, Dana Boyd Barr, Joe Brown, Jennifer Cannon, John M. Colford, Jr., Joseph Eisenberg, Ayse Ercumen, Helen Petach, Matthew Freeman, Karen Levy, Stephen P. Luby, Christine Moe, Amy J. Pickering, Jeremy A. Sarnat, Jill Stewart, Evan Thomas, Mami Taniuchi, Thomas Clasen

2.1 Context

This chapter presents a review of current approaches used to estimate human exposure to enteric pathogens. It draws from a draft manuscript titled *"Measuring environmental exposure to enteric pathogens in low-income settings: review and recommendations of an interdisciplinary working group"*, that we drafted with a an expert group of biomarker researchers, microbiologists, exposure scientists, environmental engineers, and epidemiologists after an enteric pathogen exposure workshop held at Emory University in September 2019. This workshop aimed to identify priorities for improved approaches to measuring enteric pathogen exposure by reviewing current and emerging enteric exposure assessment practices, and exploring potential lessons in exposure science from other areas of environmental health.

2.2 External and internal exposure assessments

Traditional exposure science uses numerous approaches to estimate human exposure to environmental contaminants. A recent National Research Council report, *Exposure Science in the 21st Century*, presented a summary of these approaches, ranging from those that measure environmental concentrations of contaminants to predict exposures before the contaminant reaches the human boundary, to those that estimate a dose after the contaminant has been taken up into the body (Figure 2-1).[1] As exposure assessments move beyond the human boundary they provide *internal* measurements of exposure, providing potentially more relevant information on the exposure-outcome relationship, but also fail to capture information on the *external* sources and pathways of exposure that would enable mitigation or

inform policy. For this dissertation, I am using a similar approach to defining methods used to estimate exposure to enteric pathogens, by differentiating between external and internal exposure assessments.

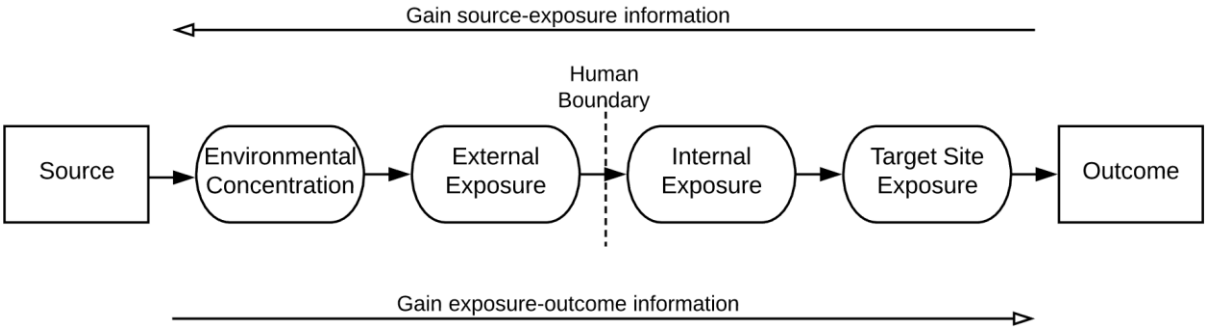


Figure 2-1: Conceptual framework for the levels of exposure assessments connecting source to outcome (adapted from [1])

External enteric exposure assessments start by estimating concentrations of enteric pathogens in different environmental reservoirs. These external measures are usually pathway-specific, an important element for enteric exposure assessments, because they can distinguish between different fecal-oral transmission pathways. However, external measures provide no information on the magnitude of total exposure to measured pathogens which must be estimated or imputed, so can only be considered as an indirect proxy for actual exposure. These surrogate measures may be more proximal to actual exposure if they include data on human interactions with their environment, such as those published by the United States Environmental Protection Agency in the *Exposure Factors Handbook*.[2] While the handbook is focused on chemical exposures in the United States, it demonstrates the types of frameworks that can be employed to inform assumptions for external exposure assessments, such as ingestion rates (drinking water, soil and food) and object mouthing.[2]

Internal enteric exposure assessments seek to estimate the actual pathogen amount that has crossed the human boundary, typically via oral ingestion. In this respect, they address the main shortcoming of the external assessment. On the other hand, they provide little information about the source or transmission pathway that external assessments offer. They can, however, provide information on the presence, types and intensities of past exposure to enteric pathogens as well as indications of potential health impacts.

While the ingested dose of enteric pathogens is not typically measured, internal exposure can be inferred from serology, detection of pathogens in feces, and other biomarkers of exposure.

2.3 Current approaches

2.3.1 Criteria

When considering different approaches to measuring exposure to enteric pathogens there are a number of criteria one might consider. These include:

- **External vs. internal**: Is exposure characterized in the environment (external) and does it provide data on the source of exposure, or is it measured after the contaminant has crossed the human boundary (internal)? For external exposure assessments, is exposure characterized proximal to the human boundary or is it a more distal measure that requires modelling to estimate more proximal exposure? For internal exposure assessments, are the measures mediated by host susceptibility to infection?
- **Pathway-specific**: Can the exposure assessment quantify the relative contribution to total exposure by different transmission pathways?
- **Granularity:** How specific is the microbiological measure (e.g. indicator of fecal contamination versus specific pathogens)? Does the method characterize presence/absence or quantitative concentrations of the contaminant? Does the assay evaluate viability or infectivity of enteric pathogens? What are the limits of detection of the assay? Is exposure to contaminants assessed at the community or individual level? Can the source of microbiological contamination (i.e., humans vs. specific animals) be ascertained? How much variability and measurement error exists with the methods?
- **Logistical considerations:** Can environmental contamination and human interaction with the environment be assessed at scale or is it constrained by cost or other factors? Are the measurement methods suitable for deployment in the field in low-resource or emergency settings? Does the assay require cold-chain transport or a consistent energy source? Are

the required materials bulky or dangerous to handle? How much training and material is required to collect samples, conduct analyses, and interpret results? Is it fast enough to provide actionable feedback to reduce exposures in a population of interest?

- **Variability:** Does the exposure assessment account for temporal and spatial variability in enteric pathogen occurrence and concentration? Does the assessment consider variability in human behaviors?
- **Ethics:** Are exposure assessment methods potentially burdensome on the communities and individuals where they are conducted? Do they require respondents to provide a substantial amount of their time and privacy? Do they provide interpretable information for end users?

2.3.2 Measuring enteric pathogens in the environment

Studies characterizing fecal exposure and some evaluations assessing the effectiveness of WaSH interventions, have mainly attempted to characterize external exposure by estimating household fecal contamination. A systematic review of the effects of sanitation interventions on fecal-oral transmission pathways identified the following approaches used: fecal pathogens or indicator bacteria in environmental samples (drinking water, hands, sentinel toys, food, household and latrine surfaces and soil); the presence or abundance of flies; and observations of human and animal feces.[3] Other research has attempted to quantify relative contributions of fecal contamination from the different transmission pathways and from private vs. public domains, including a comparison of the contributions from hands and stored water,[4] measuring fecal contamination in open drains in communities with inadequate sanitation infrastructure[5] and testing for fecal contamination and a select number of diarrheal pathogens in soil, surfaces and produce.[6] There are a number of factors to consider when measuring enteric pathogen occurrence in the environment, including environmental sampling strategies, the use of indicators as proxies for enteric pathogens, differentiating between human and animal sources of contamination, and selecting which specific pathogens to target. This section describes methods used to detect enteric pathogens in

environmental samples and provides a summary of these methods against the criteria listed above in Table 2-1.

Sampling strategies

The first decision when designing an external exposure assessment is along what fecal-oral transmission pathways to estimate enteric pathogen occurrence and where to collect samples for those pathways. For example, drinking water samples can be collected directly from the source, either at the community- or household-level, from the household storage container or from the vessel used to retrieve water for drinking from the storage container or source,[7] each progressively more proximal to the human boundary. Hand contamination can be assessed through rinsing hands in sterile water and analyzing the rinse water[8]. Hand rinse samples can be collected either directly from children or from their caregivers, each providing information on contamination along different transmission pathways. Child exposure to fecal contamination from hands can be either direct, i.e. through hand mouthing, or indirect, i.e. through hands contaminating food prepared for children.[9] Food samples can include items prepared at home or bought outside the home, for example from local markets or street food vendors.[10] Surfaces can be sampled by swabbing, and soil by scraping topsoil from a designated area.[11] The site of collection of surface swabs or soil is a key decision in sampling protocols. Sentinel toy rinses (i.e. using plastic toy balls) have been used as a proxy to characterize fecal contamination on fomites.[12] The level of contamination estimated from toy rinses depends on the level of interaction of household members with the toy. Contamination from flies has been characterized using both fly density[13] and by capturing flies and testing them for fecal indicators or specific pathogens.[14] Fly sampling can be conducted at latrine entrances to estimate fly density at a source of fecal contamination, or in food preparation areas, which is more proximal to fecal-oral transmission (i.e. through flies contaminating food).

Fecal contamination in all of these reservoirs is highly variable temporally,[15–17] seasonally,[18,19] and spatially,[20] and additional variability can be introduced by the methods used to analyze samples.[21] Some strategies to address this variability include collecting longitudinal samples over time and focusing on key

times of exposure (e.g., measuring hand cleanliness before eating) to help account for some temporal and/or seasonal variability, and selecting sampling locations as proximally as possible to the assumed true external exposure for the target population to minimize the effects of spatial variability. Timing of sample collection can also be important when estimating exposure-outcome relationships. Environmental samples are commonly collected at the same time as health data, typically for logistical convenience. Collecting samples prospectively to ascertaining health endpoints, allows for an incubation period before onset of health outcomes.[22–24] Estimating pathogen occurrence and health outcomes at the same time may be vulnerable to reverse causation (i.e. through changes in behavior due to illness such as increased drinking water treatment),[25] obscuring true exposure-outcome relationships.

Indicators of fecal contamination

Fecal contamination in the environment is commonly estimated by using indicators of fecal contamination. These indicators have the advantage that they are easier and less expensive to measure compared to multiple specific pathogens and they can be indicative of a range of enteric pathogens.[26] They provide an indication of the presence of fecal matter in a sample, but do not confirm the presence or absence of pathogens, nor do they provide any indication on the infectivity or diversity of enteric pathogens in a sample. There are a number of indicators of fecal contamination, including chemical indicators (e.g., fecal sterols, caffeine, estrogen hormones)[27,28] as well as microbial indicators such as fecal indicator bacteria (FIB). FIBs are often grouped into the coliform and streptococcal bacterial groups. Total coliforms (TC) include a broad spectrum of bacteria occurring in feces, but can also be found in non-fecal matter. Fecal coliforms (FC) or thermotolerant coliforms (TTC) are a group of bacteria that are more specific to fecal contamination, with the exception of *Klebsiella*.[29] *Escherichia coli (E. coli)* is the most commonly found FC bacteria and is more specific to human and animal fecal matter. Fecal streptococci (FS) were identified as an alternative to TC in the 1950s when it became clear that TC was a non-specific indicator for fecal contamination, but the use of this indicator diminished once methods for FC and *E.coli* culturing were established.[26] Enterococci are a subset of species of the FS group, more specific to fecal contamination than FS and more persistent in the environment.[30] Coliphages and

crAssphages,, types of bacteriophages (viruses that infect bacteria), are also used as indicators of fecal pollution because of their similar morphological characteristics and ability to mimic the persistence of viral pathogens in the environment.[31,32]

The use of fecal indicators was historically established to measure fecal contamination in drinking water and recreational waters[26] and to monitor the performance of water treatment processes. As such, fecal indicators were designed to act as process indicators and are not optimized to characterize exposure to fecal contamination in various environmental reservoirs. In the United States, TC is still used by some municipalities as a conservative measure to monitor treatment efficacy and post-treatment contamination of drinking water supplies in accordance with the Total Coliform Rule, which requires the monitoring for the presence of TC in public water systems at a frequency proportional to the number of people those systems serve.[33] *E. coli* and FC are presently the most commonly used FIB for fecal contamination monitoring in water. The World Health Organization (WHO) uses levels of *E. coli* and TTC to define microbial quality of drinking water.[34] FIBs are increasingly used to characterize fecal contamination along other fecal-oral transmission pathways. For example, enterococci is a commonly used indicator for fecal contamination on hands and fomites.[35,36]

Source tracking

FIB and some other indicators of fecal contamination do not distinguish between different sources of contamination. Fecal source tracking aims to provide data on the source of fecal contamination by detecting signatures of specific sources (e.g., particular animals or particular geographies). In addition to humans as a source of fecal loading in the environment, animal fecal contamination is of particular interest in low-income settings where households often cohabitate with animals in confined spaces. A recent systematic review outlined the importance of animals as a source of fecal contamination to human health by synthesizing evidence suggesting an association between animal feces exposure and diarrhea, soil-transmitted helminth infection, EED, growth faltering and trachoma.[37]

A variety of source tracking methods have been described in the literature, ranging from the use of a fecal coliform/fecal streptococcus ratio (because human feces are associated with higher levels of fecal coliform, whereas animal feces are associated with higher level of fecal streptococcus) to the use of host-specific molecular markers.[38] Host-specific *Bacteroidales* are commonly used in high-income countries to complement the use of fecal indicators,[39] and are seeing increased use in low-income settings.[40–42] However, fecal source tracking assays developed in one geographic location must be validated to be used in additional locations, due to geographic variation in human and animal fecal microbiomes. *Bacteroidales* can distinguish between sources of fecal contamination, because they adapt to their host differentially, allowing for the identification of host-specific fecal contamination.[43] Host-specific *Bacteroidales* have been measured in Tanzania to characterize human-specific fecal contamination on hands and in drinking water,[44] and test the relationship between human-specific fecal contamination and diarrhea.[45] A nested study of a cluster-randomized controlled sanitation trial in Odisha, India utilized fecal source tracking to discern the effectiveness of a sanitation intervention on the reduction of human-specific *Bacteroidales* compared to animal-specific *Bacteroidales.*[46]

Enteric pathogen detection in the environment

Specific pathogen occurrence or concentration, instead of the use of indicators as proxies, are less commonly measured in the environment although they may be more representative of the actual health risk associated with exposure. Pathogen detection in environmental samples is more likely to yield positive results in areas where there is high prevalence of infection with enteric pathogens and more pathogens entering the environment, but rarer pathogens will be harder to detect. Exposure assessments that measure specific pathogens need to consider not only the diversity of potential pathogens occurring in the environment but relevance of each included pathogen for health outcomes of interest, which is highlight context specific, including whether specific pathogen are associated with symptomatic or asymptomatic infections. The possibility for improved specificity from measuring specific pathogens instead of indicators of fecal contamination may come at a loss of sensitivity, since selected pathogens may not be representative of all possible pathogens in the environment.

Monitoring water supplies for pathogens can be prohibitively resource intensive due to the array of pathogens to assay for and the low concentrations of specific pathogens in the environment. The WHO estimates that concentrations of pathogens in water corresponding to 10^{-6} Disability-Adjusted Life Years (DALYs) per person per year are typically less than 1 organism per 10^4–10^5 liters.[34] Testing for pathogens in water is achieved by concentrating the water sample, such as filtering large quantities of water, on the order of 1 to 1,000 liters, followed by using culture-dependent or culture-independent methods to enumerate the occurrence and density of pathogens.[47] Filters that work for viruses may not work for protozoa and bacteria due to the size difference in pathogens, making it difficult to test for multiple pathogens simultaneously.[47] Culture-dependent methods are also limited by their low sensitivity and their resource intensity. Furthermore, some enteric pathogens (*Salmonella* Typhi, *Vibrio cholerae, Campylobacter* spp., and others) enter a viable but non-culturable state in the environment that may require special resuscitation steps or molecular methods to detect.[48]

Culture-independent molecular methods have been developed for many enteric pathogens, but they require extensive laboratory equipment and highly skilled technical staff.[49] These methods have the advantage that they can rapidly detect multiple pathogens in a sample and they circumvent the need to grow microorganisms in laboratory culture and thus can detect non-culturable organisms.[50] For example, polymerase chain reaction (PCR) based assays have become lower-cost, easier to multiplex, and more robust to inhibitors in environmental samples. A number of studies have successfully detected bacterial, viral, helminth, and protozoan pathogens in drinking water, on hands, and in soil in low-resource settings.[6,44,46,51,52] Liu *et al*. developed the customizable Taqman Array Card (TAC), an emerging method for quantitative detection of multiple enteric pathogens encompassing viral, bacterial, protozoal, and helminth targets for enteric infections.[53] Recent applications of the TAC method include the simultaneous detection of a number of enteric pathogens in surface water, soil and infant weaning food in Kenya.[54,55] However, molecular methods cannot be used to establish viability of the organism and can be prohibitively costly. Resource-intensive field collection and lab processes mean these methods are

challenging to deploy in low- and middle-income country settings where resources are limited. In addition, multi pathogen exposure assessments raise questions about the statistical analyses and interpretation of these data.[56]

Metagenomics, the sequencing and analysis of all DNA in environmental samples, circumvents the problem that many enteric pathogens cannot be easily cultured[57] and metagenomic data can provide information on the abundance and diversity of microorganisms in environmental samples. Unlike PCR methods, it does not require pre-specification of targets, allowing the user to probe for all potential enteric pathogens present in a sample. Reduced sequencing costs over time as well as recent advances in sequencing technologies and bioinformatics pipelines will continue to open up opportunities for enteric pathogen detection. Metagenomics can be used to characterize both pathogens in environmental samples and in human stool samples. Environmental metagenomics has recently been used to profile viral pathogen diversity in environmental waters,[58] and to demonstrate exchange of antibiotic resistance genes between soil bacteria and clinical pathogens.[59] Limitations of metagenomics include poor sensitivity if enteric pathogens are at low prevalence in the microbial community or when sequencing depth is low, high cost, required bioinformatics expertise, and the need for improved analysis pipelines for identifying pathogens.

Table 2-1: Summary of methods to measure enteric pathogens and indicators of fecal contamination in the environment

Tool	Source information	Proximity to outcome	Specific targets	Presence/ absence vs. concentrations	Single vs. multiple targets	Developmental stage
Environmental samples:						
Indicators (culture-based)	None	Mixed evidence on the associations between indicators and health outcomes	Some indicators include specific targets, such as E. coli and enterococci	Presence/ absence and concentration both possible	Single	Assays available and extensively used globally with increased use for environmental samples other than water
Specific pathogens						
Culture-based	None	Limited or no evidence between pathogens in environmental samples and health outcomes, except for outbreak investigations	Yes	Presence/ absence and concentration both possible	Single	Assays available but currently limited use on environmental samples in low-income settings
Molecular methods	None		Yes but need to pre-specify targets	PCR is limited to presence/ absence but density data possibly with qPCR	Single and multiple possible	
Metagenomics	None		Yes, targets do not need to be pre-specified but depend on reference library	Abundance and diversity of microbial populations, concentrations data more limited with current methods	Multiple	
Fecal source tracking						
Molecular methods	Human vs. animal	Limited or no evidence between fecal source markers in environmental samples and health outcomes	Yes	PCR is limited to presence/ absence but concentration data possibly with qPCR	Single and multiple possible	Assays available but validation work on-going to distinguish between specific animals
Metagenomics	Human vs. animal		Yes	Concentration data possible	Multiple	

2.3.3 Measuring host interaction with the environment

Exposure to enteric pathogens is not only conditional on pathogen presence in different environmental reservoirs, but also on host interaction with those reservoirs. Collecting data on host interaction with the environment is critical for exposure assessments and in the analysis of exposure data, as has been done in estimating human exposure to other environmental pollutants.[60] Survey data on self-reported behaviors or observational data on practices can enable exposure assessments in targeting environmental reservoirs and locations where the study population is predominantly exposed. Quantitative observational data can be combined with environmental measurements of enteric pathogens to estimate pathogen ingestion rates.

SaniPath, is an exposure assessment approach designed for low-income urban environments that examines exposure to fecal contamination in the public domain via ten pathways, including: drinking water, bathing water, contact with surface waters and marine waters, flood waters and open drains, ingestion of uncooked produce and street food, contact with public toilets, and soil. The SaniPath tool combines environmental sample collection and analyses for *E. coli* with surveys of behavior to better estimate exposure.[61] SaniPath highlighted the potential for microbiological and observational tools to be used in concert to estimate both fecal contamination in the environment and inform assumptions on where and how the study population is exposed. To date, the SaniPath tool has been used in ten cities in nine countries (http://sanipath.org/) to identify the major pathways of exposure to fecal contamination in the urban environment and provide evidence for urban sanitation policies and investments.

This section provides a summary of methods used to characterize human interaction with their surrounding environment and outlines methods against the criteria listed above in Table 2-2.

Surveys and self-reports

Surveys have been used as rapid and cost-effective tool to collect information on a range of self-reported behaviors that serve as proxies for exposure patterns. While surveys carry the risk of various types of bias, such as recall bias, courtesy bias, and reporting bias associated with self-report of socially desirable behaviors,[62,63] surveys are nonetheless a useful tool to obtain information on neutral behaviors that do not

trigger these biases. There are multiple approaches for collecting self-reported data. Many studies have used household surveys with the head of household or primary caregiver for young children. One major limitation of this approach is that the reported behavior for the respondent is often incorrectly seen as a proxy for behaviors for the entire household, thus risk misrepresenting exposure risk. Community participatory surveys and surveys in school classrooms, that combine some discussion of the behavior with a method for the participants in the group to confidentially report their own behavior, such as pocket voting, have also been used to identify high-risk behaviors.[61] Surveys and self-reports can also be used to inform sampling locations (e.g., where a household obtains their drinking water, prepares their food etc.).

Observations

Spot-check observations can capture WaSH infrastructure and behaviors that result in risk of exposure (e.g., latrine cleanliness, presence of a handwashing station, handwashing at key moments, washing raw produce before consumption) that can be difficult to elicit by self-report due to biased reporting. More structured observations,[64,65] including the use of videography,[66,67] offer an opportunity to gather information on complex behaviors – including recording the frequency, duration, and type of interaction with the environment, which could subsequently be used to estimate ingestion rates. However, they are resource intensive at scale and suffer from high heterogeneity of within- and between-host behavior.[68] There are questions about whether this heterogeneity may introduce prohibitive amounts of uncertainty into ingestion rate estimates. Observations can also cause reactivity in participants where the presence of an outside observer leads individuals to alter their behaviors while observed.[69] Sensors have been used to compare observed behaviors to reported behaviors with some studies indicating that reported behaviors are inconsistent with sensor measured use.[70–73] Sanitary surveys, i.e. survey-based inspections of water systems, have been designed and promoted, often in combination with periodic water quality testing, as screening and risk assessment tools for fecal contamination exposure. However, recent studies have shown a poor correlation between sanitary inspection scores and actual fecal contamination in drinking water supplies.[74]

Location tracking

Personal global positioning system (GPS) tracking can inform where a host is spending time and can thus provide data on where to collect environmental samples, for example by identifying potential hot spots in communities where community members may be experiencing frequent exposures. The use of GPS devices has broadly been validated for exposure assessments,[75] and has been used to inform air pollution[76] and chemical exposure assessments.[77] A study in Brazil found that GPS tracking was an effective tool to quantify personal movements of urban slum residents and evaluate exposure sources of environmental leptospirosis transmission.[78]

Tracers

Tracers, substances introduced into the environment so that their distribution can be detected from their distinctive properties, can provide data on where and how hosts are interacting with their environment. They have been used in air pollution epidemiology to differentiate between indoor and outdoor exposures[79] and a study in China estimated child soil ingestion by measuring concentrations of tracer elements in soil.[80] Challenges with using tracers include that seeding several common fecal-oral transmission pathways simultaneously to quantify relative exposure contributions from different pathways may be impractical beyond certain microenvironments, and tracers are needed that do not degrade in the environment and pose no risk to human or environmental health.

Table 2-2: Summary of methods to characterize interaction with the environment

Tool	Source information	Pathway incrimination	Granularity (Community vs. individual)	Risk of bias
Surveys	Limited	Limited	Community and individual possible	High
Self-reports	Limited	Limited	Individual	High
Structured/unstructured observations	Possible	Possible	Individual	High
Video observations	Possible	Possible	Individual	Medium
GPS tracking	None	No	Individual	Low
Tracers	None	Theoretically possible but no data available on tracers being implemented in practice.	Individual	Low

2.3.4 Measuring enteric pathogens in humans

Measuring internal exposure to enteric pathogens can provide data on past exposure, although findings from these measures are complicated by internal human biological mechanisms and host immunity. These data can be employed to evaluate how well interventions reduce exposure to specific pathogens and which pathogens humans are exposed to in their community. Data generated for epidemiological studies and for surveillance has benefited from advances in enteric pathogen detection methodologies to estimate the burden of disease, disease severity, and attributing health outcomes to pathogens and pathogens to exposure pathways.[73] These data facilitate hazard characterization, the first step of risk assessment. Exposure assessments can take into account the hazard characterization information derived from epidemiologic studies to focus on pathogens that cause the greatest disease burden in the region of interest, taking a more narrowed approach to exposure assessments, for example by focusing on specific pathogen-source pairs. This section describes methods used to detect enteric pathogens in human biological samples and provides a summary of these methods against the criteria listed above in Table 2-3.

Pathogen shedding in stool

Methods to detect enteric pathogens in stool samples range from using microscopy[82] or enzyme-linked immunosorbent assays[83] to detect single pathogens, to using molecular or metagenomics methods to characterize multiple pathogens in a sample. Multiplex PCR is a technique that has been widely employed in enteric disease surveillance (sporadic and outbreak) and epidemiological studies such as in multi-country case-control[84] and longitudinal birth cohort studies,[85] as well as in recent studies measuring health impacts of WaSH interventions.[86,87] Metagenomic approaches have also been employed, for example to distinguish between foodborne disease outbreak strains of *Salmonella*,[88], and to identify the likely causes of diarrheagenic *E. coli* in Ecuador.[89] The advantage of these internal measures providing enteric infection data is that they provide evidence of actual ingestion of enteric pathogens. The limitation of these techniques is not only the intensive resources they require, but it can also be difficult to attribute a specific pathogen to a disease outcome when multiple enteric pathogens are detected in stool

simultaneously or when asymptomatic infections are common. Additionally, the sensitivity of these methods can vary, so it is critical to define upper and lower limits of detection. These methods detect infections, and pathogens may only be shed by an infected person for a short period of time (days, weeks) or shed intermittently, so infections between sampling events would be missed. Furthermore, the duration of shedding after infection is highly pathogen-specific, so these methods can be biased towards persistent pathogens that shed for a longer period of time compared to more transient pathogens.

Pathogen-specific immunoassays

Another way of estimating past exposure to enteric pathogens is through immunological assays detecting pathogen-specific antibodies in serum or saliva, which can be multiplexed to detect exposure to multiple enteric pathogens.[90] For these immunological methods, the timeline of exposure can be difficult to ascribe as low levels of pathogen-specific antibodies (Immunoglobulin (Ig)A, and particularly IgG) can be present in saliva and serum for weeks to years after infection.[91] This can also be an advantage, as the methods can be used to integrate prior exposure over longer periods of time, rather than relying on pathogen shedding in stool. Exposure data without regard to history of infection can be useful for some applications, such as to determine if a population has been exposed to a rare or emerging pathogen or particular microbial strain, which could be important for focusing exposure assessment approaches. Sero-epidemiology is a promising approach to measurement of force of infection of enteropathogens across entire populations.[92–94] However, one's immune response depends on a number of host-specific factors including history of previous exposure (acquired immunity[95]), nutritional status,[96] genetics,[97] composition of the gut microbiome,[98] underlying disease such as HIV infection,[99] and age (antibodies can appear in low concentrations in young children, particularly in saliva).[100]

Table 2-3: Summary of methods to estimate internal exposure

Tool	Source information	Proximity to outcome	Specific targets	Presence/ absence vs. concentrations	Single vs. multiple targets	Developmental stage
Human samples:						
Stool-based assays:	None	Measure of infection as a proxy for internal exposure More proximal to outcomes of interest than environmental measures				
Microscopy			Yes, but typically limited to protozoan pathogens and soil transmitted helminths (STH)	Presence/ absence	Single and multiple possible for protozoan pathogens and STHs	Commonly used as a well-established method
Culture-based			Yes	Presence/ absence and concentration both possible	Single	Assays available for different pathogens
Molecular methods			Yes	PCR is limited to presence/ absence but concentration data possibly with qPCR	Single and multiple possible	Assays recently used in large scale cohort studies and intervention evaluations in low-income settings
Metagenomics			Yes	Concentration data possible	Multiple	Limited data on metagenomics approaches used for pathogen detection in stool in low-income settings
Enzyme-linked immunosorbent assays			Yes	Concentration data possible but difficult to compare	Single and multiple possible	Assays available to detect immunological responses in serum and saliva, and for antigen detection in stool

2.4 Limitations of current approaches

The approaches presented both for external and internal measures of exposure highlight that these methods are limited to providing proxy estimates of exposure, i.e. there is no gold standard approach to estimating external exposure to enteric pathogens from various fecal-oral transmission pathways or to quantify the ingested dose. One of the primary limitations of external measures of exposure is from the use of fecal indicators to estimate fecal contamination at the household level. While the measurement of fecal indicators may be valuable as an indication of the fecal load in environmental samples and to provide a common metric for comparison with previous studies, the shortcomings of indicators when they are used for purposes other than their original intended use of monitoring water supply systems, have been well documented,[101] including reports of *E. coli* detection in pristine areas of tropical and even temperate environments.[102–107] If measures of external exposure include survey and observational tools, exposure assessments can move from the household-level to being more proximal to estimating individual-level exposures, for example by providing data to estimate individual ingestion rates. However, challenges associated with observational methods include characterizing the variability of human interaction with different environmental reservoirs, so the heterogeneity of data output from these methods may introduce prohibitive amounts of uncertainty into ingestion rate estimates.

Findings from recent evaluations WaSH interventions highlight the limitations of using household-level estimates of fecal indicators. Luby and colleagues reported protective effects on diarrhea from the tested WaSH interventions except water treatment with chlorine, even though reductions in FIB were found in both stored drinking water and food,[108] suggesting that the chlorine-susceptible bacteria used to measure fecal contamination were not representative of chlorine-resistant pathogens, such as protozoa, contributing to waterborne disease. This is consistent with the trial's findings of reductions in *Giardia* infection in all WaSH arms but the water treatment in households from this arm.[109] The same evaluation also found protective effects on diarrhea and infections with protozoa and soil-transmitted helminths in the sanitation-only arm,[109,110] even though there was no evidence of a change in FIB in water, food, soil,

on hands or a change in fly density,[111] suggesting reductions in disease transmission not captured by the FIB measurements. Reese and colleagues reported no impact on diarrhea but reduced stunting from a water supply and sanitation intervention despite no evidence of a reduction in fecal contamination of drinking water or hands.[112] Pickering and colleagues also found no effect of a community-led sanitation intervention on diarrhea but an improvement in child growth, despite no reduction in fecal contamination in drinking water; however, latrine fly presence and observed human and animal feces did significantly decrease in the treatment group.[113]

The shortcomings of using fecal indicators are amplified by the limited association between indicators and the presence of enteric pathogens.[114,115] Indicators have been commonly used for environmental samples in part because of the challenges associated with detecting specific pathogens using traditional approaches, such as the need for large sample volumes, different detection methods depending on the pathogen and resource-intensive field collection and lab processes. As a result, detecting specific pathogens in environmental samples can be challenging in low- and middle-income country settings where resources are limited, contamination with multiple different pathogens is common and household water can be a scarce commodity. However, recent advances in environmental metagenomics[57] and multiplex molecular methods[53] mean that detecting multiple specific pathogens in environmental samples may soon be possible at scale.

Using various human biological matrices to characterize internal exposure, such as stool, serum or saliva, does not estimate the ingested dose directly but rather provides an indication of past exposure. Furthermore, even when specific pathogens are not detected in biological samples, a host may still have been exposed to those pathogens. There is limited evidence on how much of an ingested dose reaches the target cells and is capable of inducing infection. Animal models often do not exist for enteric pathogens or do not cause the same health outcomes as in humans. Dose-response data are limited because human challenge studies are difficult, expensive, usually single-pathogen focused, primarily performed among adults in high-income countries, cannot control for previous exposures and differences in immune

responses to infection, and may be at odds with acceptable ethical standards.[116] In addition, one's immune response depends on a number of host-specific factors including history of previous exposure (acquired immunity),[95] nutritional status,[96] genetics,[97] composition of the gut microbiome,[98] underlying disease such as HIV infection,[99] and age (antibodies can appear in low concentrations in young children, particularly in saliva).[100] For stool samples, pathogens may only be shed by an infected person for a short period of time (days, weeks) or shed intermittently,[117] making collection of relevant stool samples more complex. It can also be difficult to attribute a specific pathogen to a disease outcome when multiple enteric pathogens are detected in stool simultaneously or when asymptomatic infections are common.

2.5 Lessons from other areas of environmental health

2.5.1 Measurement error

Analytical frameworks have been introduced to assess the impact of exposure measurement error or the effect of using proxy measures of external exposure in health effects models. There is now a large body of research on measurement error introduced by assigning exposure using stationary outdoor air-monitoring sites, a method that is used widely to establish regulatory limits on ambient air quality,[118] as surrogates of actual individual-level exposures.[119] This research has elucidated the presence and effects introduced by these proxy measures within many types of epidemiologic models. Although the nature of measurement error when estimating exposure to environmental contaminants may be complex and multifactorial, studies have shown that some forms of error may lead to biases and greater uncertainties in estimating a true exposure-outcome relationship.[120]

2.5.2 Air quality

In air pollution epidemiology, moving from population-level measures of exposure to measures that better reflect exposure at the individual level, and thus reduce exposure measurement error, has been a focus of research for the past two decades. For example, in 1999, the National Research Council Committee on *Research Priorities for Airborne Particulate Matter* outlined the quantification of the difference between proxy and personal measures of exposure as a key research priority for better understanding differences in

observed health risk estimates across individuals and locations.[121] Since then, new approaches have been designed to measure an individual's inhalation exposure to particulate matter in air, using personal breathing zone samplers. To characterize exposures to cookstove emissions, in particular, new devices, such as the Enhanced Children's MicroPEM (ECM)[122] and the Ultrasonic Personal Air Sampler (UPAS),[123] have been developed to provide more precise measurements for use in public health research applications in low-income countries. Personal exposure monitors have not been developed to estimate exposure to enteric pathogens, perhaps in part because of some of the differences between characterizing exposure to particulate matter and pathogens. These include, that pathogens can be transmitted via multiple pathways and the challenges associated with real-time detection of a potentially diverse set of pathogens found in low concentrations in the environment.

Measuring long-term exposures to air pollution, a known driver for a range of chronic adverse health effects, necessitates alternative approaches for characterizing exposure, and often employ hybrid methods, which combine both modeling and personal monitoring. A promising, and increasingly common approach for estimating spatiotemporally-resolved long-term exposures to particulate matter and several gaseous pollutants comes from satellite remote sensing.[124] These methods have the ability to use satellite optical instrumentation, calibrated with ground-level ambient monitoring data, to create long-term global exposure surfaces.[125] Some water quality parameters can be measured via remote sensing (e.g., chlorophyll-a, nutrient concentrations), but these methods apply more to large water bodies than to individual glasses of water consumed by humans.[126] Other methods integrate human activity patterns, questionnaires related to sources of exposure, and actual measurements conducted within defined settings (i.e., microenvironments) to predict individual level air pollution exposures over short- and long-term periods. The Air Pollution Exposure Model (APEX) is an example of this class of air pollution exposure model, which was developed in response to prior limitations related to air pollution exposure and which may offer insights for novel, combined approaches for characterizing exposures to enteric pathogens as

well.[60] Similarly, there are analogous hydrological water quality models that have been developed for estimating exposures to water pollution for surface water bodies.

2.5.3 Chemical Toxicants

Chemical toxicant exposure assessments, the measurement of a chemical, or its metabolite, degradate, reaction product or surrogate, include external and internal exposure assessments. External chemical exposure assessments can be pathway-specific (e.g., water, air) and route-specific (e.g., ingestion, inhalation) while internal assessments integrate all pathways and routes of exposure through which a chemical has entered the body. For example, assessment of dermal chemical exposures includes the use of hand wipes, patches and body suits as dosimeters of exposure. The dosimeters are removed after exposure and chemical concentrations are measured in them, estimating dermal exposure.[127] Similarly, personal air space pumps or patches are used to estimate inhalational exposures and duplicate diet or food or water measurements may be used to estimate ingestion exposures.[2] Emerging techniques such as the use of silicon wristbands to absorb airborne contaminants have also been used as efficient means to capture exposure to up to 150 contaminants in air, including polychlorinated biphenyls, pesticides, flame retardants, polycyclic aromatic hydrocarbons and volatile organic chemicals.[128] Advances in exposomics using high-resolution metabolomics have created advanced methods of internal exposure assessments with the simultaneous detection of numerous endogenous and exogenous chemical metabolites in human biological samples.[129] Biomonitoring is also seeing application in air pollution exposure assessments in low-income settings, for example with the detection of polycyclic aromatic hydrocarbons in urine to quantify household air pollution exposure.[130]

2.6 References

2.7 1References

1 The National Research Council. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC, 2012 DOI:10.17226/13507.

2 US EPA. Exposure Factors Handbook: 2011 Edition. 2011 DOI:EPA/600/R-090/052F.

3 Sclar GD, Penakalapati G, Amato HK, *et al.* Assessing the impact of sanitation on indicators of fecal exposure along principal transmission pathways: A systematic review. *Int J Hyg Environ Health* 2016; **219**: 709–23.

4 Mattioli MCM, Davis J, Boehm AB. Hand-to-mouth contacts result in greater ingestion of feces than dietary water consumption in Tanzania: A quantitative fecal exposure assessment model. *Environ Sci Technol* 2015; **49**: 1912–20.

5 Gretsch SR, Ampofo JA, Baker KK, *et al.* Quantification of exposure to fecal contamination in open drains in four neighborhoods in Accra, Ghana. *J Water Health* 2016; **14**: 255–66.

6 Pickering AJ, Julian TR, Marks SJ, *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.

7 Rufener S, Mäusezahl D, Mosler H-J, Weingartner R. Quality of drinking-water at source and point-of-consumption--drinking cup as a high potential recontamination risk: a field study in Bolivia. *J Health Popul Nutr* 2010; **28**: 34–41.

8 Pickering AJ, Boehm AB, Mwanjali M, Davis J. Efficacy of waterless hand hygiene compared with handwashing with soap: A field study in Dar es Salaam, Tanzania. *Am J Trop Med Hyg* 2010; **82**: 270–8.

9 Julian TR. Environmental transmission of diarrheal pathogens in low and middle income countries. Environ. Sci. Process. Impacts. 2016; **18**: 944–55.

10 Lanata CF. Studies of food hygiene and diarrhoeal disease. *Int J Environ Health Res* 2003; **13**: 175–83.

11 Steinbaum L, Njenga SM, Kihara J, *et al.* Soil-Transmitted Helminth Eggs Are Present in Soil at Multiple Locations within Households in Rural Kenya. *PLoS One* 2016; **11**: e0157780.

12 Vujcic J, Ram PK, Hussain F, *et al.* Toys and toilets: Cross-sectional study using children's toys to evaluate environmental faecal contamination in rural Bangladeshi households with different sanitation facilities and practices. *Trop Med Int Heal* 2014; **19**: 528–36.

13 Bell M, Irish S, Schmidt WP, Nayak S, Clasen T, Cameron M. Comparing trap designs and methods for assessing density of synanthropic flies in Odisha, India. *Parasites and Vectors* 2019; **12**. DOI:10.1186/s13071-019-3324-z.

14 Doza S, Rahman MJ, Islam MA, *et al.* Prevalence and association of Escherichia coli and diarrheagenic Escherichia coli in stored foods for young children and flies caught in the same households in Rural Bangladesh. *Am J Trop Med Hyg* 2018; **98**: 1031–8.

15 Levy K, Hubbard AE, Nelson KL, Eisenberg JNS. Drivers of Water Quality Variability in Northern Coastal Ecuador. *Environ Sci Technol* 2009; **43**: 1788–97.

16 Ram PK, Jahid I, Halder AK, *et al.* Variability in hand contamination based on serial measurements: implications for assessment of hand-cleansing behavior and disease risk. *Am J*

Trop Med Hyg 2011; **84**: 510–6.

17 Pickering AJ, Julian TR, Mamuya S, Boehm AB, Davis J. Bacterial hand contamination among Tanzanian mothers varies temporally and following household activities. *Trop Med Int Heal* 2011; **16**: 233–9.

18 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.

19 Ercumen A, Pickering AJ, Kwong LH, *et al.* Animal Feces Contribute to Domestic Fecal Contamination: Evidence from E. coli Measured in Water, Hands, Food, Flies, and Soil in Bangladesh. *Environ Sci Technol* 2017; **51**: 8725–34.

20 Rao G, Eisenberg JNS, Kleinbaum DG, Cevallos W, Trueba G, Levy K. Spatial variability of Escherichia coli in rivers of northern coastal Ecuador. *Water* 2015; **7**: 818–32.

21 Levy K, Nelson KL, Hubbard A, Eisenberg JNS. Rethinking Indicators of Microbial Drinking Water Quality for Health Studies in Tropical Developing Countries: Case Study in Northern Coastal Ecuador. *Am J Trop Med Hyg* 2012; **86**: 499–507.

22 Ercumen A, Arnold BF, Naser AM, Unicomb L, Colford JM, Luby SP. Potential sources of bias in the use of Escherichia coli to measure waterborne diarrhoea risk in low-income settings. *Trop Med Int Heal* 2017; **22**: 2–11.

23 Pickering AJ, Ercumen A, Arnold BF, *et al.* Fecal Indicator Bacteria along Multiple Environmental Transmission Pathways (Water, Hands, Food, Soil, Flies) and Subsequent Child Diarrhea in Rural Bangladesh. *Environ Sci Technol* 2018; **52**: 7928–36.

24 Luby SP, Halder AK, Huda TM, *et al.* Microbiological Contamination of Drinking Water Associated with Subsequent Child Diarrhea. *Am J Trop Med Hyg* 2015; **93**: 904–11.

25 Khush RS, Arnold BF, Srikanth P, *et al.* H2S as an Indicator of Water Supply Vulnerability and Health Risk in Low-Resource Settings: A Prospective Cohort Study. *Am J Trop Med Hyg* 2013; **89**: 251–9.

26 The National Research Council. Indicators for Waterborne Pathogens. Washington, DC: National Academies Press (US), 2004 https://www.ncbi.nlm.nih.gov/books/NBK215658/ (accessed June 12, 2018).

27 Sankararamakrishnan N, Guo Q. Chemical tracers as indicator of human fecal coliforms at storm water outfalls. *Environ Int* 2005; **31**: 1133–40.

28 Srijew S, Chiemchaisri W, Chiemchaisri C, Satoh H. Source identification of fecal contamination in the canals in Bangkok using fecal sterol compounds. *J Water Sanit Hyg Dev* 2016; **6**: 13–22.

29 US EPA. Water: Monitoring and Assessment - Fecal Bacteria. https://archive.epa.gov/water/archive/web/html/vms511.html (accessed June 12, 2018).

30 Sinton LW, Donnison AM, Hastie CM. Faecal streptococci as faecal pollution indicators: A review. Part II: Sanitary significance, survival, and use. *New Zeal J Mar Freshw Res* 1993; **27**: 117–37.

31 McMinn BR, Ashbolt NJ, Korajkic A. Bacteriophages as indicators of faecal pollution and enteric virus removal. Lett. Appl. Microbiol. 2017; **65**: 11–26.

32 Dutilh BE, Cassman N, Mcnair K, *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* 2014; **5**. DOI:10.1038/ncomms5498.

33 US EPA. 40 CFR Parts 141 and 142 National Primary Drinking Water Regulations: Revisions to

the Total Coliform Rule. 2013 www.regulations.gov (accessed Dec 3, 2019).

34 WHO. Guidelines for Drinking-water Quality: 4th Edition. Geneva, 2017 http://apps.who.int/iris/bitstream/handle/10665/254637/9789241549950-eng.pdf?sequence=1 (accessed June 12, 2018).

35 Julian TR, Pickering AJ, Leckie JO, Boehm AB. Enterococcus spp on fomites and hands indicate increased risk of respiratory illness in child care centers. *Am J Infect Control* 2013; **41**: 728–33.

36 JUDAH G, DONACHIE P, COBB E, SCHMIDT W, HOLLAND M, CURTIS V. Dirty hands: bacteria of faecal origin on commuters' hands. *Epidemiol Infect* 2010; **138**: 409.

37 Penakalapati G, Swarthout J, Delahoy MJ, *et al.* Exposure to Animal Feces and Human Health: A Systematic Review and Proposed Research Priorities. *Environ Sci Technol* 2017; **51**: 11537–52.

38 Scott TM, Rose JB, Jenkins TM, Farrah SR, Lukasik J. Microbial Source Tracking: Current Methodology and Future Directions. *Appl Environ Microbiol* 2002; **68**: 5796–803.

39 Harwood VJ, Staley C, Badgley BD, Borges K, Korajkic A. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiol Rev* 2014; **38**: 1–40.

40 Boehm AB, Wang D, Ercumen A, *et al.* Occurrence of Host-Associated Fecal Markers on Child Hands, Household Soil, and Drinking Water in Rural Bangladeshi Households. *Environ Sci Technol Lett* 2016; **3**: 393–8.

41 Harris AR, Pickering AJ, Harris M, *et al.* Ruminants Contribute Fecal Contamination to the Urban Household Environment in Dhaka, Bangladesh. *Environ Sci Technol* 2016; **50**: 4642–9.

42 Harris AR, Islam MA, Unicomb L, *et al.* Fecal contamination on produce from wholesale and retail food Markets in Dhaka, Bangladesh. *Am J Trop Med Hyg* 2018; **98**: 287–94.

43 Bernhard AE, Field KG. A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. *Appl Environ Microbiol* 2000; **66**: 4571–4.

44 Mattioli MC, Pickering AJ, Gilsdorf RJ, Davis J, Boehm AB. Hands and Water as Vectors of Diarrheal Pathogens in Bagamoyo, Tanzania. *Environ Sci Technol* 2013; **47**: 355–63.

45 Mattioli MC, Boehm AB, Davis J, Harris AR, Mrisho M, Pickering AJ. Enteric pathogens in stored drinking water and on caregiver's hands in Tanzanian households with and without reported cases of child diarrhea. *PLoS One* 2014; **9**: e84939.

46 Odagiri M, Schriewer A, Daniels ME, *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.

47 Straub TM, Chandler DP. Towards a unified system for detecting waterborne pathogens. *J Microbiol Methods* 2003; **53**: 185–97.

48 Ramamurthy T, Ghosh A, Pazhani GP, Shinoda S. Current perspectives on viable but non-culturable (VBNC) pathogenic bacteria. Front. Public Heal. 2014; **2**. DOI:10.3389/fpubh.2014.00103.

49 Ramírez-Castillo FY, Loera-Muro A, Jacques M, *et al.* Waterborne pathogens: detection methods and challenges. *Pathogens* 2015; **4**: 307–34.

50 Douterelo I, Boxall JB, Deines P, Sekar R, Fish KE, Biggs CA. Methodological approaches for studying the microbial ecology of drinking water distribution systems. *Water Res* 2014; **65**: 134–56.

51 Fuhrmeister ER, Ercumen A, Pickering AJ, *et al.* Predictors of Enteric Pathogens in the Domestic Environment from Human and Animal Sources in Rural Bangladesh. *Environ Sci Technol* 2019; **53**: 10023–33.

52 Amoah ID, Singh G, Stenström TA, Reddy P. Detection and quantification of soil-transmitted helminths in environmental samples: A review of current state-of-the-art and future perspectives. Acta Trop. 2017; **169**: 187–201.

53 Liu J, Gratz J, Amour C, *et al.* A laboratory-developed taqman array card for simultaneous detection of 19 enteropathogens. *J Clin Microbiol* 2013; **51**: 472–80.

54 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**: 10263–74.

55 Tsai K, Simiyu S, Mumma J, *et al.* Enteric Pathogen Diversity in Infant Foods in Low-Income Neighborhoods of Kisumu, Kenya. *Int J Environ Res Public Health* 2019; **16**: 506.

56 Medgyesi D, Sewell D, Senesac R, Cumming O, Mumma J, Baker KK. The landscape of enteric pathogen exposure of young children in public domains of low-income, urban Kenya: The influence of exposure pathway and spatial range of play on multi-pathogen exposure risks. *PLoS Negl Trop Dis* 2018; **13**. DOI:10.1371/journal.pntd.0007292.

57 Tringe SG, Rubin EM. Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* 2005; **6**: 805–14.

58 McCall C, Xagoraraki I. Metagenomic Approaches for Detecting Viral Diversity in Water Environments. *J Environ Eng (United States)* 2019; **145**. DOI:10.1061/(ASCE)EE.1943-7870.0001548.

59 Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. The shared antibiotic resistome of soil bacteria and human pathogens. *Science (80-)* 2012; **337**: 1107–11.

60 Özkaynak H, Baxter LK, Dionisio KL, Burke J. Air pollution exposure prediction approaches used in air pollution epidemiology studies. *J Expo Sci Environ Epidemiol* 2013; **23**: 566–72.

61 Robb K, Null C, Teunis P, Yakubu H, Armah G, Moe CL. Assessment of Fecal Exposure Pathways in Low-Income Urban Neighborhoods in Accra, Ghana: Rationale, Design, Methods, and Key Findings of the SaniPath Study. *Am J Trop Med Hyg* 2017; **97**: 1020–32.

62 Manun'Ebo M, Cousens S, Haggerty P, Kalengaie M, Ashworth A, Kirkwood B. Measuring hygiene practices: a comparison of questionnaires with direct observations in rural Zaire. *Trop Med Int Heal* 1997; **2**: 1015–21.

63 Stanton BF, Clemens JD, Aziz KM, Rahman M. Twenty-four-hour recall, knowledge-attitude-practice questionnaires, and direct observations of sanitary practices: a comparative study. *Bull World Health Organ* 1987; **65**: 217–22.

64 Kwong LH, Ercumen A, Pickering AJ, Unicomb L, Davis J, Luby SP. Hand- and object-mouthing of rural bangladeshi children 3–18 months old. *Int J Environ Res Public Health* 2016; **13**. DOI:10.3390/ijerph13060563.

65 Kwong LH, Ercumen A, Pickering AJ, Unicomb L, Davis J, Luby SP. Age-related changes to environmental exposure: variation in the frequency that young children place hands and objects in their mouths. *J Expo Sci Environ Epidemiol* 2019; : 1.

66 Julian TR, Pickering AJ. A pilot study on integrating videography and environmental microbial sampling to model fecal bacterial exposures in peri-urban Tanzania. *PLoS One* 2015; **10**: 1–15.

67 Pickering AJ, Blum AG, Breiman RF, Ram PK, Davis J. Video surveillance captures student hand hygiene behavior, reactivity to observation, and peer influence in Kenyan primary schools. *PLoS One* 2014; **9**: 1–7.

68 Teunis PFM, Reese HE, Null C, Yakubu H, Moe CL. Quantifying Contact with the Environment: Behaviors of Young Children in Accra, Ghana. *Am J Trop Med Hyg* 2016; **94**: 920–31.

69 Ram PK, Halder AK, Granger SP, *et al.* Is Structured Observation a Valid Technique to Measure Handwashing Behavior? Use of Acceleration Sensors Embedded in Soap to Assess Reactivity to Structured Observation. *Am J Trop Med Hyg* 2010; **83**: 1070–6.

70 Delea MG, Nagel CL, Thomas EA, *et al.* Comparison of respondent-reported and sensor-recorded latrine utilization measures in rural Bangladesh: A cross-sectional study. *Trans R Soc Trop Med Hyg* 2017; **111**: 308–15.

71 Bohnert K, Chard AN, Mwaki A, *et al.* Comparing sanitation delivery modalities in urban informal settlement schools: A randomized trial in Nairobi, Kenya. *Int J Environ Res Public Health* 2016; **13**. DOI:10.3390/ijerph13121189.

72 Sinha A, Nagel CL, Thomas E, *et al.* Assessing latrine use in rural India: A cross-sectional study comparing reported use and passive latrine use monitors. *Am J Trop Med Hyg* 2016; **95**: 720–7.

73 Thomas EA, Tellez-Sanchez S, Wick C, *et al.* Behavioral Reactivity Associated With Electronic Monitoring of Environmental Health Interventions A Cluster Randomized Trial with Water Filters and Cookstoves. *Environ Sci Technol* 2016; **50**: 3773–80.

74 Snoad C, Nagel C, Bhattacharya A, Thomas E. The Effectiveness of Sanitary Inspections as a Risk Assessment Tool for Thermotolerant Coliform Bacteria Contamination of Rural Drinking Water: A Review of Data from West Bengal, India. *Am J Trop Med Hyg* 2017; **96**: 976–83.

75 Beekhuizen J, Kromhout H, Huss A, Vermeulen R. Performance of GPS-devices for environmental exposure assessment. *J Expo Sci Environ Epidemiol* 2013; **23**: 498–505.

76 Breen MS, Long TC, Schultz BD, *et al.* GPS-based microenvironment tracker (MicroTrac) model to estimate time-location of individuals for air pollution exposure assessments: model evaluation in central North Carolina. *J Expo Sci Environ Epidemiol* 2014; **24**: 412–20.

77 Rohlman D, Dixon HM, Kincl L, *et al.* Development of an environmental health tool linking chemical exposures, physical location and lung function. *BMC Public Health* 2019; **19**: 854.

78 Owers KA, Odetunde J, de Matos RB, *et al.* Fine-scale GPS tracking to quantify human movement patterns and exposure to leptospires in the urban slum environment. *PLoS Negl Trop Dis* 2018; **12**: e0006752.

79 Brinkman GL, Milford JB, Schauer JJ, Shafer MM, Hannigan MP. Source identification of personal exposure to fine particulate matter using organic tracers. *Atmos Environ* 2009; **43**: 1972–81.

80 Lin C, Wang B, Cui X, *et al.* Estimates of Soil Ingestion in a Population of Chinese Children. *Environ Health Perspect* 2017; **125**. DOI:10.1289/EHP930.

81 Dowell SF, Blazes D, Desmond-Hellmann S. Four steps to precision public health. Nature. 2016; **540**: 189–91.

82 Garcia LS, Arrowood M, Kokoskin E, *et al.* Laboratory diagnosis of parasites from the gastrointestinal tract. *Clin Microbiol Rev* 2018; **31**: e00025-17.

83 Ghoshal U, Jain V, Dey A, Ranjan P. Evaluation of enzyme linked immunosorbent assay for stool antigen detection for the diagnosis of cryptosporidiosis among HIV negative

immunocompromised patients in a tertiary care hospital of northern India. *J Infect Public Health* 2018; **11**: 115–9.

84 Liu J, Platts-Mills JA, Juma J, *et al.* Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *Lancet* 2016; **388**: 1291–301.

85 Platts-Mills JA, Babji S, Bodhidatta L, *et al.* Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet Glob Heal* 2015; **3**: e564–75.

86 Brown J, Cumming O, Bartram J, *et al.* A controlled, before-and-after trial of an urban sanitation intervention to reduce enteric infections in children: research protocol for the Maputo Sanitation (MapSan) study, Mozambique. *BMJ Open* 2015; **5**. DOI:10.1136/bmjopen-2015-008215.

87 Arnold BF, Null C, Luby SP, *et al.* Cluster-randomised controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: the WASH Benefits study design and rationale. *BMJ Open* 2013; **3**. DOI:10.1136/bmjopen-2013-003476.

88 Huang AD, Luo C, Pena-Gonzalez A, Weigand MR, Tarr CL, Konstantinidis KT. Metagenomics of two severe foodborne outbreaks provides diagnostic signatures and signs of coinfection not attainable by traditional methods. *Appl Environ Microbiol* 2017; **83**: e02577-16.

89 Peña-Gonzalez A, Soto-Girón MJ, Smith S, *et al.* Metagenomic Signatures of Gut Infections Caused by Different Escherichia coli Pathotypes. *Appl Environ Microbiol* 2019; **85**: e01820-19.

90 Exum NG, Pisanic N, Granger; DA, *et al.* Use of Pathogen-Specific Antibody Biomarkers to Estimate Waterborne Infections in Population-Based Settings. *Curr Environ Heal Reports* 2016; **3**: 322–34.

91 Zabriskie JB. Essential Clinical Immunology. Cambridge, UK: Cambridge University Press, 2009 http://sacema.org/uploads/Essential-Clinical-Immunology.pdf (accessed July 5, 2019).

92 Arnold BF, Martin DL, Juma J, *et al.* Enteropathogen antibody dynamics and force of infection among children in low-resource settings. *Elife* 2019; **8**: e45594.

93 Arnold BF, Scobie HM, Priest JW, Lammie PJ. Integrated serologic surveillance of population immunity and disease transmission. *Emerg Infect Dis* 2018; **24**: 1188–94.

94 Monge S, Teunis P, Friesema I, *et al.* Immune response-eliciting exposure to Campylobacter vastly exceeds the incidence of clinically overt campylobacteriosis but is associated with similar risk factors: A nationwide serosurvey in the Netherlands. *J Infect* 2018; **77**: 171–7.

95 Swift L, Hunter PR. What do negative associations between potential risk factors and illness in analytical epidemiological studies of infectious disease really mean? *Eur J Epidemiol* 2003; **19**: 219–23.

96 Watson RR, McMurray DN, Martin P, Reyes MA. Effect of age, malnutrition and renutrition on free secretory component and IgA in secretions. *Am J Clin Nutr* 1985; **42**: 281–8.

97 Scepanovic P, Alanio C, Hammer C, *et al.* Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med* 2018; **10**. DOI:10.1186/s13073-018-0568-8.

98 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**: ofx144.

99 Chowdhury A, Silvestri G. Host-pathogen interaction in HIV infection. Curr. Opin. Immunol. 2013; **25**: 463–9.

100 Brandtzaeg P. Secretory immunity with special reference to the oral cavity. *J Oral Microbiol* 2013; **5**. DOI:10.3402/jom.v5i0.20401.

101 Leclerc H, Mossel DAA, Edberg SC, Struijk CB. Advances in the Bacteriology of the Coliform Group: Their Suitability as Markers of Microbial Water Safety. *Annu Rev Microbiol* 2001; **55**: 201–34.

102 Oh S, Buddenborg S, Yoder-Himes DR, Tiedje JM, Konstantinidis KT. Genomic Diversity of Escherichia Isolates from Diverse Habitats. *PLoS One* 2012; **7**: e47005.

103 Byappanahalli M, Fujioka R. Indigenous soil bacteria and low moisture may limit but allow faecal bacteria to multiply and become a minor population in tropical soils. *Water Sci Technol* 2004; **50**: 27–32.

104 Rivera SC, Hazen TC, Toranzos GA. Isolation of fecal coliforms from pristine sites in a tropical rain forest. *Appl Environ Microbiol* 1988; **54**: 513–7.

105 Field KG, Samadpour M. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res. 2007; **41**: 3517–38.

106 Solo-Gabriele HM, Wolfert MA, Desmarais TR, Palmer CJ. Sources of Escherichia coli in a coastal subtropical environment. *Appl Environ Microbiol* 2000; **66**: 230–7.

107 Ishii S, Ksoll WB, Hicks RE, *et al.* Presence and Growth of Naturalized Escherichia coli in Temperate Soils from Lake Superior Watersheds Presence and Growth of Naturalized Escherichia coli in Temperate Soils from Lake Superior Watersheds. 2006; **72**: 612–21.

108 Ercumen A, Mertens A, Arnold BF, *et al.* Effects of Single and Combined Water, Sanitation and Handwashing Interventions on Fecal Contamination in the Domestic Environment: A Cluster-Randomized Controlled Trial in Rural Bangladesh. *Environ Sci Technol* 2018; **52**: 12078–88.

109 Lin A, Ercumen A, Benjamin-Chung J, *et al.* Effects of Water, Sanitation, Handwashing, and Nutritional Interventions on Child Enteric Protozoan Infections in Rural Bangladesh: A Cluster-Randomized Controlled Trial. *Clinincal Infect Dis* 2018; **67**: 1515–22.

110 Ercumen A, Benjamin-Chung J, Arnold BF, *et al.* Effects of water, sanitation, handwashing and nutritional interventions on soil-transmitted helminth infections in young children: A cluster-randomized controlled trial in rural Bangladesh. *PLoS Negl Trop Dis* 2019; **13**: e0007323.

111 Ercumen A, Pickering AJ, Kwong LH, *et al.* Do Sanitation Improvements Reduce Fecal Contamination of Water, Hands, Food, Soil, and Flies? Evidence from a Cluster-Randomized Controlled Trial in Rural Bangladesh. *Environ Sci Technol* 2018; **52**: 12089–97.

112 Reese HE, Brown J, Pickering AJ, Mishra SR, Chang HH, Clasen TF. The effect of a combined household piped-water and sanitation intervention on fecal contamination of source water, drinking water, and children's hands: a matched cohort study in Odisha, India. .

113 Pickering AJ, Djebbari H, Lopez C, Coulibaly M, Alzua ML. 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.

114 Wu J, Long SC, Das D, Dorner SM. Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. *J Water Health* 2011; **9**: 265–78.

115 Hörman A, Rimhanen-finne R, Maunula L, Bonsdorff C Von, Torvela N. Indicator Organisms in Surface Water in and Indicator Organisms in Surface Water in Southwestern. *Appl Environ*

Microbiol 2004; **70**: 87–95.

116 Bambery B, Selgelid M, Weijer C, Savulescu J, Pollard AJ. Ethical criteria for human challenge studies in infectious diseases. *Public Health Ethics* 2016; **9**: 92–103.

117 European Centre for Disease Prevention and Control. Systematic review on the incubation and infectiousness/ shedding period of communicable diseases in children. Stockholm, 2016 DOI:10.2900/337026.

118 US EPA. Reviewing National Ambient Air Quality Standards (NAAQS): Scientific and Technical Information. https://www.epa.gov/naaqs (accessed Aug 7, 2019).

119 Weisskopf MG, Webster TF. Trade-offs of Personal Versus More Proxy Exposure Measures in Environmental Epidemiology. *Epidemiology* 2017; **28**: 635–43.

120 Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010; **340**: c2289.

121 The National Research Council. Research Priorities for Airborne Particulate Matter: II. Evaluating Research Progress and Updating the Portfolio. Washington, DC: The National Academies Press, 1999 DOI:10.17226/9646.

122 Burrowes VJ, Johnson M, Piedrahita R, *et al.* Pilot Study of the Enhanced Children's Micropem Compared to the Standard Gravimetric Methods to Measure Personal Exposure and Kitchen Concentrations to Fine Particulate Matter from Household Air Pollution: HAPIN Trial Investigators. In: ISEE Conference Abstracts. 2018. DOI:10.1289/isesisee.2018.P01.0460.

123 Volckens J, Quinn C, Leith D, Mehaffy J, Henry CS, Miller-Lionberg D. Development and evaluation of an ultrasonic personal aerosol sampler. *Indoor Air* 2017; **27**: 409–16.

124 van Donkelaar A, Martin R V., Brauer M, *et al.* Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ Health Perspect* 2010; **118**: 847–55.

125 Larkin A, Hystad P. Towards Personal Exposures: How Technology Is Changing Air Pollution and Health Research. Curr. Environ. Heal. reports. 2017; **4**: 463–71.

126 Gholizadeh MH, Melesse AM, Reddi L. A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors. 2016; **16**. DOI:10.3390/s16081298.

127 Semple S. Dermal exposure to chemicals in the workplace: just how important is skin absorption? *Occup Environ Med* 2004; **61**: 376–82.

128 Anderson KA, Points GL, Donald CE, *et al.* Preparation and performance features of wristband samplers and considerations for chemical exposure assessment. *J Expo Sci Environ Epidemiol* 2017; **27**: 551–9.

129 Lankadurai BP, Nagato EG, Simpson MJ. Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. *Environ Rev* 2013; **21**: 180–205.

130 Li Z, Commodore A, Hartinger S, *et al.* Biomonitoring Human Exposure to Household Air Pollution and Association with Self-reported Health Symptoms-A Stove Intervention Study in Peru HHS Public Access. *Environ Int* 2016; **97**: 195–203.

3 Chapter 3 – Research Aim 1: Fecal contamination of the environment and child health: A systematic review and meta-analysis using individual participant data

Manuscript authors: Frederick G.B. Goddard, Amy J. Pickering, Ayse Ercumen, Joe Brown, Howard H. Chang, Thomas Clasen

3.1 Context

The mixed impact of WaSH interventions on child health outcomes could indicate that interventions did not sufficiently reduce exposure to fecal contamination, or that household fecal contamination as a proxy for individual-level exposure to enteric pathogens is not sufficiently associated with adverse child health outcomes. For the first research aim of this dissertation we built a database of household-level fecal contamination data along common fecal-oral transmission pathways as proxies for external enteric exposure, and matched those data to child health outcomes. These data were collected under the umbrella of WaSH intervention evaluations, and we employed them here to estimate enteric exposure-health outcome relationships.

3.2 Background

Traditional WaSH approaches have focused on reducing open defecation, promoting improved sanitation, encouraging handwashing with soap and improving the quality of and access to water. While reducing fecal contamination and subsequent occurrence of enteric pathogens in the household environment is an essential condition for WaSH interventions to improve infectious disease burdens, few studies actually measure the association between fecal contamination and health or the effects of interventions on fecal exposure. Those that do are largely confined to measurements of drinking water quality, one of multiple possible sources of exposure. Pooled analyses of these studies have been limited to quantifying the relationship between fecal contamination in drinking water and diarrhea, and have yielded differing results. Gundry et al. reported no association between three FIB and diarrhea (OR: 1.12, 95% CI 0.85, 1.48);[1] Gruber et al. found a significant association between levels of *E. coli* in drinking water and

diarrhea (RR: 1.54, 95% CI 1.37, 1.74) but not for fecal coliforms (RR: 1.07, 95% CI 0.79, 1.45);[2] Hodge et al., the only study using individual participant data (IPD), showed a higher odds of diarrhea with increasing levels of fecal coliforms in drinking water (OR: 1.18, 95% CI 1.11, 1.26), but was limited to seven studies.[3]

A study by Pickering, Ercumen et al. in rural Bangladesh investigated the relationship between household fecal contamination along multiple pathways (water, soil, food, hands and flies) and WHO- and caregiver defined diarrhea as well as bloody stool, measured concurrently and prospectively.[4] Results from this study were mixed, finding evidence of an association between *E. coli* levels on child hands and WHO- and caregiver defined diarrhea (IRR 1.23, 95% CI 1.06, 1.43; IRR 1.31, 95% CI 1.11, 1.55) and between *E. coli* in food and bloody stool (IRR 1.34, 95% CI 1.07, 1.68) when samples were collected prospectively. This study also found an association between *E. coli* levels in flies and soil and caregiver-defined diarrhea (PR 1.15, 95% CI 1.04, 1.26; PR 1.16, 95% CI 1.02, 1.32) when samples were collected concurrently with diarrhea data.

Mixed results from these studies and findings from recent WaSH evaluations have raised questions about the relative contribution of different fecal-oral transmission pathways to adverse acute enteric health outcomes, such as diarrhea, and the relationship between fecal contamination and child linear growth. Our aim was to conduct a systematic review and meta-analysis using IPD to examine the relationship between fecal contamination and child health. Specifically, we sought to test whether fecal contamination along common transmission pathways, as measured by FIB and fly densities, is associated with diarrhea and linear growth in children under the age of five in low- and middle-income countries. Unlike a conventional meta-analysis that pools available estimates of effect from eligible studies, the use of IPD offered the opportunity to implement consistent analytical approaches across individual studies, as well as assess these relationships between child health and fecal contamination along common exposure pathways even when individual studies did not publish the effect estimates of interest.[5]

3.3 Methods

3.3.1 Search strategy and selection criteria

We searched the PubMed, Web of Science and Embase databases using the search strings specified in Appendix A. The search was conducted on May 21, 2018 and included all study designs. We included published studies, studies with a published protocol and studies identified from conference abstracts. During the initial title/abstract screening, studies were included for further review if they measured diarrheal disease prevalence or child linear growth. Studies were excluded if they 1) were set in high-income countries; 2) were conducted in the public domain (i.e. schools, hospitals, child care centers); 3) did not include data on children under the age of five; or 4) pertained to reviews, commentaries, progress reports or studies that used large national demographic survey data (i.e. studies that conducted no original data collection). For the diarrhea analysis, all diarrhea definitions were eligible but due to risk of recall bias studies were excluded if they used a recall period longer than seven days.[6] For the linear growth analyses, cross-sectional studies were excluded, due to the lack of longitudinal environmental contamination data measured prior to the growth outcomes. Full texts included after title/abstract review were reviewed against the same inclusion and exclusion criteria. In addition, studies were excluded during full text review if they did not measure FIB concentrations along at least one of the major fecal-oral transmissions pathways. FIB concentrations are less commonly measured fFor flies, so we included studies that measured fly density in with fly traps in food preparation areas over a period of time (typically 24 hours) as a proxy for the transmission of fecal contamination on to household food.

Title/abstract screening and full text reviews were duplicated by FG and a research assistant. The search was conducted in English but full texts published in French, Portuguese and Spanish were also reviewed. For studies determined to be eligible after full text review, we sought IPD from corresponding authors and extracted relevant summary information from the manuscripts. Studies for which IPD were not available were not included in a parallel conventional meta-analyses, because studies typically did not report fecal exposure-health outcome associations. This study was registered with PROSPERO

(CRD42018102114) and the study protocol with a pre-specified analysis plan was made publicly available on the Open Science Framework (OSF) (https://osf.io/G6AKD/) before the beginning of data analysis. Updates to the search strategy and analysis plan since publication of the original protocol are also publicly available on the OSF.

3.3.2 Data analysis

We requested caregiver-reported diarrhea and height/length measurements at the individual level, along with child age, gender and survey date. At the household level we requested FIB concentrations in drinking water, soil, food, fomite and hand rinses, as well as fly densities in food preparation areas, treatment status and environmental sample collection date. These reflect the expected common pathways of fecal exposure.[7] As clustering variables, we requested unique identifiers for each child, household and community to allow for adjustment of clustered health outcomes. These data were anonymized prior to analysis. We did not require additional institutional review board approval for use of these data. We extracted information on study location, study design, FIB used and whether study communities were urban or rural from study protocols or manuscripts. To control for the effects of precipitation on exposure,[8] our model included a term differentiating between wet and dry season months based on the 30-year average monthly precipitation for each included study from the WorldClim dataset.[9]

We then matched household-level fecal contamination data to individual-level health data for all children under the age of five. For the diarrhea analyses, single time-point environmental samples collected on the same day or up to seven days before diarrhea data collection were matched to caregiver-reported diarrhea. For the linear growth analyses, we matched all available environmental samples collected during a child's life up to the day anthropometric measurements were taken and calculated the median. We transformed FIB concentrations and fly densities into categorical variables based on a $\log_{10}$ scale with four levels of contamination as our exposure variables: <1, 1-10, 11-100 and >100 colony-forming unit (CFU) or most probable number (MPN) for FIB (per 100ml for hand and fomite rinses, and drinking water; per dry g for food; per dry mg for soil) and <1, 1-10, 11-100 and >100 flies per 24 hrs for fly densities. We generated

height-for-age Z-scores (HAZ) with the height/length, age and gender data using the World Health Organization (WHO) growth standards.[10] For precipitation, we classified a month as a wet season month if average precipitation was >60mm, and as a dry season month if average precipitation was <60mm, based on the Köppen–Geiger climate classification system.[11]

We conducted our analyses using multilevel generalized mixed effects models. Primary outcomes were diarrhea and HAZ scores, and parameters of interest were the odds ratio (OR) for diarrhea and change in HAZ scores associated with a 1-$\log_{10}$ difference in measures of fecal contamination along different transmission pathways, indexed by *p* below. We also included the odds of stunting (HAZ score less than -2) as a secondary outcome (not pre-specified). We modeled each study individually using the following models for the primary outcomes:

$$logit\left(d_{p,ijk}\right) = u_{ijk} + u_{jk} + u_k + \beta_1 FC_{p,jk} + \beta_2 Age_{ijk} + \beta_3 Treat_{jk} + \beta_4 Resid_k + \beta_5 Season_k$$

$$haz_{p,ijk} = u_{jk} + u_k + \beta_1 FC_{p,jk} + \beta_2 Age_{ijk} + \beta_3 Treat_k + \beta_4 Resid_k$$

FC represents the $\log_{10}$ categories of fecal contamination (0, 1, 2 and 3), based on a single time-point sample in the diarrhea model and based on the median of longitudinal samples in the linear growth model. We controlled for child age (in years), treatment status (any intervention versus no intervention) and residence (urban versus rural). For the diarrhea model we also controlled for season (wet versus dry). In studies that had multiple children within households or communities, we controlled for clustering at household-level *j* and community-level *k*. For the diarrhea model we also controlled for clustering at child-level *i* in longitudinal studies that had repeated matched individual-level diarrhea reports and environmental samples. We combined effect estimates from each study in a meta regression using a random effects model to account for between-study heterogeneity. We characterized between-study heterogeneity using I^2 to describe the percentage of total variation across studies that is due to heterogeneity rather than chance.[12]

We conducted subgroup analyses for pathways with sufficient data by stratifying the analyses by child age, by each $\log_{10}$ category compared to no contamination, treatment status, FIB type and urban versus rural communities. Age stratification was based on the US EPA *Guidance for Selecting Age Groups for Monitoring and Assessing Childhood Exposures to Environmental Contaminants*: Birth to <3 months, 3 to <6 months, 6 to <12 months, 1 to <2 years and 2 to <5 years.[13] Stratification by $\log_{10}$ category and FIB type was not pre-specified in our analysis plan and was added in order to test how FIB used modified the exposure-outcome associations and to compare effect sizes for different levels of fecal contamination to no measured contamination. For the diarrhea analyses we also stratified by wet versus dry season and timing of environmental and diarrhea data collection. We stratified timing of data collection by cross-sectional (diarrhea and environmental data collected on the same day) versus prospective (diarrhea data collected 1-7 days after environmental data) data collection. We used the same statistical analysis methodology for all subgroup analyses as we did for the main outcomes, except for age and $\log_{10}$ category stratification where we pooled from all studies in the same model due to data sparsity and controlled for clustering at the study and community levels. To consider the effects of transforming the exposure variables in to four categories of fecal contamination, we conducted sensitivity analyses that both expanded to six $\log_{10}$ categories (i.e. <1, 1-10, 11-100, 101-1000, 1001-10,000, and +10,000) and used a continuous $\log_{10}$ transformation. All analyses were completed in R version 3.6.[14]

We assessed the risk of bias for the participating studies for each outcome and fecal-oral transmission pathway separately. We used a modified version of the Liverpool Quality Assessment Tool (LQAT), an adaptation of the Newcastle-Ottowa scale.[15] We chose the LQAT because it was adaptable to different study designs and considers the risk of bias in both exposure and outcome measures.[16] We assessed risk of selection, response rate and follow-up bias by examining how households were chosen for environmental sample collection and the integrity of collected environmental data. We also assessed risk of bias from data collection methods for exposure and outcome assessment, as well as from blinding.

3.4 Results

We screened 2,318 studies and sought IPD from 73 eligible studies (Figure 3-1). We received IPD from 30 studies, 10 of which did not meet all eligibility criteria after review of the data. Twenty studies were included in the meta-analyses,[17–36] of which all had diarrhea data and seven had linear growth data. Of the 42 studies we were not able to acquire IPD for, only six estimated fecal contamination along pathways other than drinking water and eight measured linear growth. Tables 3-1 and 3-2 present a summary of study characteristics and outlines what fecal-oral transmission pathways studies provided data for. We matched health outcome observations (diarrhea or linear growth) to fecal contamination data in drinking water (N = 54,225), followed by child hand rinses (N = 10,732), kitchen fly densities (N = 10,514) and fomites (N = 5,913). Fecal contamination on fomites was characterized using sentinel toy rinses. Toys are a fomite young children readily interact with and thus is used to characterize fomite fecal contamination.[37] For food and soil we only received data from one study, so we were not able to conduct pooled meta-analyses. We received data from countries in South America, Sub-Saharan Africa, South and South-East Asia. Most of the data originated from rural settings and FIB used were *E. coli* and fecal coliforms, both of which are designated by the WHO as indicators for fecal contamination.[38]

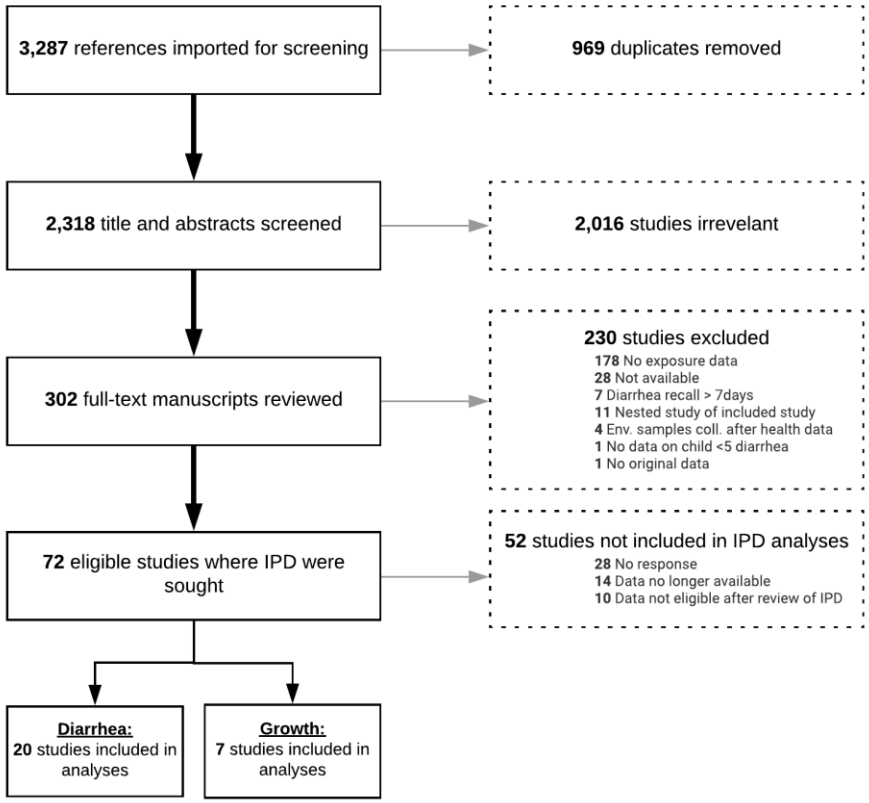


Figure 3-1: Study selection

Table 3-1: Study summary characteristics for the diarrhea analyses

Study ID	Study Design	Interv. study	Country	Urban/ Rural	FIB used	Data availability by pathway[1]						Diarrhea definition	Recall period (days)
						W	H	Fom	FI	F	S		
Arnold, 2010	Cohort	Yes	India	Rural	E. Coli	X						WHO-defined[2] or 1 or more stools with blood in 24 hours	7
Benjamin-Chung, 2018	RCT	No	Bangladesh	Rural	E. Coli	X		X	X			WHO-defined or 1 or more stools with blood in 24 hours	7
Boisson, 2010	RCT	No[3]	DRC	Rural	TTC	X						WHO-defined	7
Boisson, 2013	RCT	Yes	India	Rural & Urban	TTC	X						WHO-defined	3
Brown, 2008	RCT	Yes	Cambodia	Rural	E. Coli	X						WHO-defined or 1 or more stools with blood in 24 hours	7
Clasen, 2005	RCT	Yes	Colombia	Rural	TTC	X						WHO-defined	7
Clasen, 2014	RCT	Yes	India	Rural	TTC	X			X			WHO-defined	7
Davis, in prep.	RCT	Yes	Tanzania	Rural & Urban	E. Coli	X						WHO-defined	2
Devamani, 2014	Cross-sectional	Yes	Mozambique	Urban	E.Coli		X					WHO-defined or 1 or more stools with blood in 24 hours	7
Ercumen 2015	RCT	Yes	Bangladesh	Rural	E. Coli	X						WHO-defined	7
Kirby, 2017	Cohort	Yes	Rwanda	Rural	TTC	X						WHO-defined	7
Kirby, Nagel, 2019	RCT	Yes	Rwanda	Rural	TTC	X						WHO-defined	7
Luby, 2015	Cohort	Yes	Bangladesh	Rural	E. Coli	X						WHO-defined	2
Patil, 2015	RCT	Yes	India	Rural	E. Coli	X						WHO-defined or 1 or more stools with blood in 24 hours	7
Peletz, 2011		No	Zambia	Rural	TTC	X						WHO-defined	7
Peletz, 2012	RCT	Yes	Zambia	Peri-Urban[4]	TTC	X						WHO-defined	7
Pickering, Ercumen, 2018	RCT	Yes	Bangladesh	Rural	E. Coli	X	X		X	X	X	WHO-defined	7
Pickering, 2019	RCT	Yes	Kenya	Rural	E. Coli	X	X	X	X			WHO-defined	7
Reese, 2019	Cohort	Yes	India	Rural	E. Coli	X	X					WHO-defined	7
Sinharoy, 2017	RCT	Yes	Rwanda	Rural	TTC	X						WHO-defined	7

[1]W = drinking water; H = child hands; Fom = Fomites; FI = fly density in food preparation area; F = Food; S = Soil
[3]Only data at baseline used (no intervention implemented), no usable exposure data during trial
[2]Three or more loose or watery stools in a 24 hour period
[4]Classified as urban in our analysis

Table 3-2: Study summary characteristics for the linear growth analyses

Study ID	Study Design	Interv. study	Country	Urban/ Rural	FIB used	Data availability by pathway[1] W	H	Fom	FI	F	S
Arnold, 2010	Cohort	Yes	India	Rural	E. Coli	X					
Clasen, 2014	RCT	Yes	India	Rural	TTC	X			X		
Patil, 2015	RCT	Yes	India	Rural	E. Coli	X					
Pickering, Ercumen, 2018	RCT	Yes	Bangladesh	Rural	E. Coli	X	X	X	X	X	X
Pickering, 2019	RCT	Yes	Kenya	Rural	E. Coli	X	X	X	X		
Reese, 2019	Cohort	Yes	India	Rural	E. Coli	X	X				
Sinharoy, 2017	RCT	Yes	Rwanda	Rural	TTC	X					

[1]W = drinking water; H = child hands; Fom = Fomites; FI = fly density in food preparation area; F = Food; S = Soil

We found higher odds of diarrhea with 1-$\log_{10}$ higher FIB concentrations in drinking water (OR: 1.09; 95% CI 1.04, 1.13; p = 0.0002; I^2 = 34%) and on child hands (OR 1.11; 95% CI 1.02, 1.22; p = 0.021; I^2 = 0%). There was no evidence that FIB on fomites (OR 1.05; 95% CI 0.94, 1.16; p = 0.40; I^2 = 0%) or food preparation area fly density (OR 0.95 95% CI 0.82, 1.11; p = 0.54; I^2 = 10%) was associated with diarrhea (Figure 3-2).

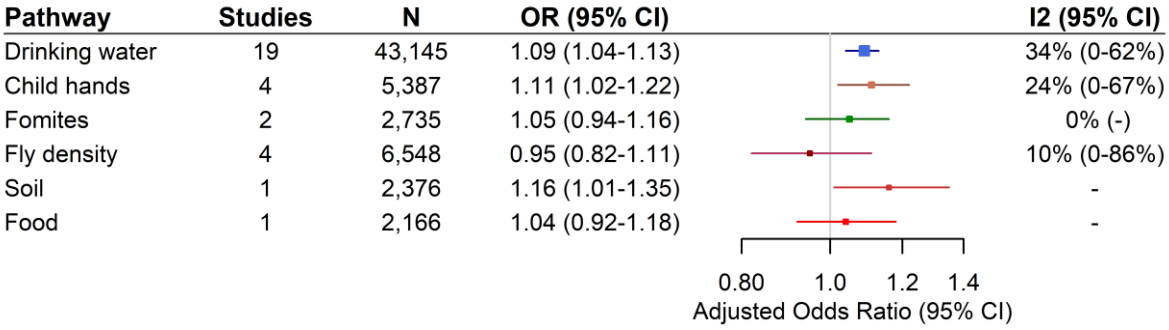


Figure 3-2: Odds of diarrhea for 1-$\log_{10}$ higher fecal contamination by fecal-oral transmission pathway

For the linear growth analyses, matched median fecal contamination was derived from 1-8 samples, depending on data availability. We found lower HAZ scores with 1-$\log_{10}$ higher median FIB concentrations in drinking water (HAZ -0.04; 95% CI -0.06, -0.01; p = 0.0054; I^2 = 19%) and on fomites (HAZ -0.06; 95% CI -0.12, 0.00; p = 0.044; I^2 = 57%). There was no evidence that FIB on child hands

(HAZ -0.02; 95% CI -0.09, 0.04; p = 0.49; I^2 = 61%) or food preparation area fly density (HAZ 0.03; 95% CI -0.04, 0.11; p = 0.41; I^2 = 26%) was associated with linear growth (Figure 3-3).

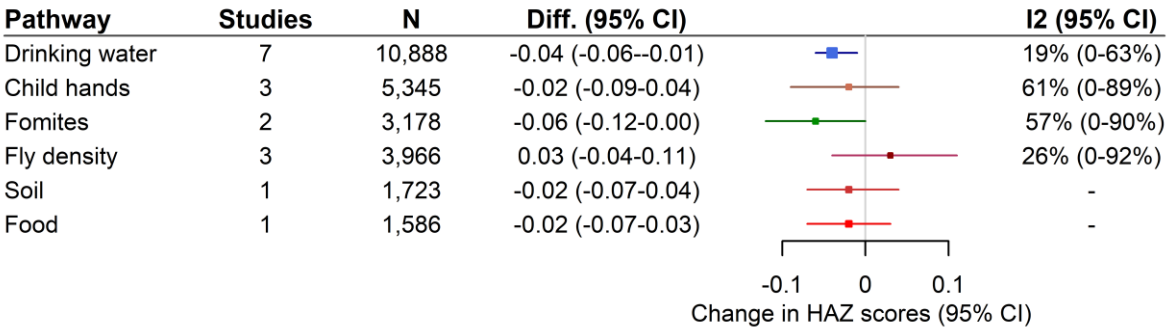


Figure 3-3: Difference in HAZ score for a median 1-log_{10} higher fecal contamination by fecal-oral transmission pathway

For child stunting, we found higher odds of stunting with 1-log_{10} higher median FIB concentrations on child hands (OR: 1.08; 95% CI 1.02, 1.15; p = 0.0055; I^2 = 0%) and fomites (OR 1.10; 95% CI 1.03, 1.19; p = 0.0086; I^2 = 0%). There was weaker evidence for an association between median FIB concentrations in drinking water and stunting (OR 1.05; 95% CI 1.00, 1.11; p = 0.062; I^2 = 30%) (Appendix B, Figure B-1). Sensitivity analyses suggested similar findings for both the diarrhea and linear growth analyses after using different transformations of exposure variables using both six log_{10} categories and continuous log_{10} (Figures B2-B5). Forest plots detailing the meta-analyses for diarrhea and linear growth and stunting with single-study findings are provided in Appendix B, Figures B-6, B-7 and B-8.

Secondary analyses that stratified by level of fecal contamination suggested that the odds of diarrhea with FIB concentrations in drinking water and on child hands, and the lower HAZ scores with FIB concentrations in drinking water and on fomites was driven by high levels of measured contamination above 10 CFU/MPN per 100ml sample and was less evident for 1-10 CFU/MPN (Figures B-9 and B-10). Age-stratified analyses of the pooled data suggested that the higher odds of diarrhea associated with higher FIB concentrations in drinking water affected all children except those under six months (Figure B-11). For child hand and fomite rinses, a higher odds of diarrhea was observed for children aged 12-24

months. For child linear growth, we found lower HAZ scores for children aged 6-24 months and >12 months with higher median FIB levels in drinking water and on fomites, respectively (Figure B-12).

Other stratified analyses suggested that there may be a stronger relationship between fecal contamination in water and diarrhea in urban compared to rural communities (Urban: OR 1.25; 95%CI 1.12, 1.40, Rural: OR 1.07; 95%CI 1.02, 1.11) (Tables 3-3 and 3-4). For diarrhea, we found no other differences by treatment status, FIB used, urban versus rural communities, dry versus wet season and prospective versus cross-sectional diarrhea data collection. Other stratified analyses for linear growth showed no differences between strata.

Table 3-3: Results from the stratified diarrhea analyses

Stratification	Pathway					
	Drinking water	**Child hands**	**Fomites**	**Flies**	**Food**	**Soil**
Log_{10} category (CFU/MPN or fly density)						
0	Reference	Reference	Reference	Reference	Reference	Reference
1-10	1.00 (0.90, 1.11)	1.05 (0.81, 1.35)	0.88 (0.68, 1.14)	0.95 (0.81, 1.12)	1.12 (0.84, 1.53)	1.10 (0.59, 2..06)
11-100	1.08 (0.98, 1.19)	1.39 (1.03, 1.89)	0.95 (0.69, 1.30)	0.91 (0.68, 1.21)	1.08 (0.75, 1.55)	1.12 (0.67, 1.87)
+100	1.30 (1.19, 1.43)	1.43 (0.99, 2.09)	1.12 (0.79, 1.59)	0.57 (0.13, 2.54)	1.17 (0.80, 1.70)	1.43 (0.91, 2.25)
Age (months)						
0-3	0.99 (0.84, 1.16)	1.22 (0.97, 1.53)	No data available	0.70 (0.23, 2.11)	1.03 (0.82, 1.28)	0.99 (0.78, 1.24)
3-6	0.95 (0.84, 1.07)	1.05 (0.86, 1.27)	No data available	0.83 (0.43, 1.62)	0.99 (0.84, 1.18)	1.24 (1.02, 1.51)
6-12	1.12 (1.03, 1.21)	1.05 (0.88, 1.27)	0.96 (0.75, 1.23)	0.92 (0.67, 1.26)	1.08 (0.26, 4.40)	1.32 (0.17, 9.96)
12-24	1.09 (1.04, 1.15)	1.38 (1.15, 1.65)	1.18 (1.00, 1.23)	0.96 (0.80, 1.15)	0.38 (0.07, 2.13)	0.65 (0.31, 1.37)
24-60	1.10 (1.06, 1.15)	0.94 (0.79, 1.13)	0.93 (0.79, 1.10)	0.98 (0.81, 1.19)	1.42 (0.83, 2.43)	2.07 (0.81, 5.29)
Treatment Status						
Intervention	1.11 (1.05, 1.18)	1.10 (0.95, 1.27)	0.92 (0.77, 1.11)	0.93 (0.74, 1.13)	1.06 (0.93, 1.22)	1.18 (1.01, 1.38)
Control	1.05 (0.99, 1.11)	1.14 (0.99, 1.31)	1.11 (0.95, 1.31)	0.98 (0.75, 1.28)	1.00 (0.83, 1.20)	1.09 (0.89, 1.33)
Fecal Indicator Bacteria						
E. Coli	1.06 (0.99, 1.15)	All E. Coli	All E. Coli	Not applicable	All E. Coli	All E. Coli
TTC	1.09 (1.04, 1.13)					
Residence						
Urban	1.25 (1.12, 1.40)	1.20 (0.78, 1.82)	All rural	All rural	All rural	All rural
Rural	1.07 (1.02, 1.11)	1.11 (1.01, 1.22)				
Season						
Dry	1.06 (0.99, 1.13)	1.14 (0.98, 1.34)	1.03 (0.78, 1.34)	0.84 (0.35, 2.05)	1.04 (0.89, 1.21)	1.11 (0.96, 1.29)
Wet	1.10 (1.04, 1.17)	1.10 (0.99, 1.22)	1.05 (0.94, 1.18)	1.00 (0.84, 1.20)	1.03 (0.87, 1.22)	1.29 (0.98, 1.70)
Sample Collection						
Prospective	1.10 (1.01, 1.19)	1.11 (1.00, 1.22)	1.06 (0.93, 1.20)	1.03 (0.89, 1.18)	1.00 (0.85,1.15)	1.28 (1.07, 1.53)
Cross-sectional	1.08 (1.02, 1.14)	1.13 (0.92, 1.39)	1.02 (0.84, 1.24)	0.79 (0.60, 1.06)	1.09 (0.93, 1.27)	1.00 (0.82, 1.20)

Table 3-4: Results from the stratified linear growth analyses

	Pathway					
Stratification	**Drinking water**	**Child hands**	**Fomites**	**Flies**	**Food**	**Soil**
Log category (CFU/MPN or fly density)						
0	Reference	Reference	Reference	Reference	Reference	Reference
1-10	-0.05 (-0.11, 0.02)	0.01 (-0.11, 0.13)	-0.06 (-0.17, 0.05)	0.06 (-0.02, 0.14)	-0.20 (-0.34, -0.07)	0.33 (0.09, 0.57)
11-100	-0.06 (-0.12, 0.01)	-0.02 (-0.12, 0.09)	-0.19 (-0.30, -0.08)	0.05 (-0.01, 0.19)	-0.03 (-0.17, 0.12)	0.11 (-0.10, 0.31)
+100	-0.09 (-0.16, -0.02)	-0.03 (-0.13, 0.07)	-0.23 (-0.34, -0.12)	0.24 (-0.16, 0.64)	-0.07 (-0.22, 0.09)	0.07 (-0.13, 0.26)
Age (months)						
0-3	0.00 (-0.15, 0.15)	No data available	No data available	No data available	No data available	No data available
3-6	-0.02 (-0.14, 0.10)	-0.20 (-0.44, 0.04)	No data available	-0.27 (-0.95, 0.41)	No data available	No data available
6-12	-0.05 (-0.09, -0.01)	-0.03 (-0.08, 0.02)	No data available	-0.16 (-0.39, 0.06)	-0.14 (-0.42, 0.14)	-0.13 (-0.44, 0.19)
12-24	-0.04 (-0.07, -0.01)	-0.02 (-0.05, 0.01)	-0.07 (-0.11, -0.03)	0.12 (0.05, 0.19)	-0.03 (-0.09, 0.02)	-0.04 (-0.10, 0.02)
24-60	-0.02 (-0.06, 0.01)	-0.02 (-0.08, 0.04)	-0.07 (-0.13, -0.01)	0.07 (-0.02, 0.15)	0.03 (-0.07, 0.13)	0.08 (-0.05, 0.22)
Treatment Status						
Intervention	-0.02 (-0.07, 0.02)	-0.02 (-0.07, 0.02)	-0.07 (-0.12, -0.02)	0.00 (-0.08, 0.07)	-0.02 (-0.08, 0.04)	0.02 (-0.04, 0.09)
Control	-0.05 (-0.09, -0.01)	-0.03 (-0.14, 0.08)	-0.05 (-0.15, 0.05)	0.04 (-0.07, 0.14)	-0.02 (-0.10, 0.07)	-0.10 (-0.19, -0.01)
Fecal Indicator Bacteria						
E. Coli	-0.02 (-0.07, 0.02)	All E. Coli	All E. Coli	Not applicable	All E. Coli	All E. Coli
TTC	-0.05 (-0.09, -0.01)					
Residence						
Urban	All rural	All rural	All rural	All rural	All rural	All rural
Rural						

Findings from our risk of bias assessment suggest that the greatest risk of bias for both the diarrhea and linear growth analyses originates from the outcome assessments, in large part because it is not possible to blind respondents or data collection staff from most WaSH interventions (Appendix C). This may introduce bias in to the analyses, for example if interventions reduced levels of household fecal contamination and members of those households were also less likely to report diarrhea. This may introduce bias in to the analyses, for example if interventions affected levels of household fecal contamination and members of those households were also less likely to report diarrhea. Our analyses also included studies that carried possible risks of selection, response rate and follow-up bias, primarily where environmental assessments were tertiary outcomes, so study design was not tailored towards optimal environmental sample collection.

3.5 Discussion

In contrast to mixed results from individual studies and meta-analyses using summary estimates, this analysis of nearly ninety thousand individual participant data points from 20 studies suggests that household FIB concentrations are associated with child diarrhea and impaired linear growth. Evidence of associations varied by exposure pathway as did data availability, with FIB concentrations in drinking water representing over half of the included environmental data, whereas food and soil were sampled in only one study. In our analyses we were not able to control for all potential confounders, so our findings should be interpreted with caution. We also found low to moderate between-study heterogeneity in some meta-analyses. These results are the first to pool data from a number of settings to find that fecal contamination in the household environment is associated with impaired child linear growth, and that fecal contamination along pathways other than drinking water can have adverse effects on child health. The findings support a recent consensus statement by WaSH researchers recommending that interventions focus more closely on the need to reduce fecal contamination in the domestic environment to achieve consistent child health benefits.[39]

In our age-stratified analyses, we found that interventions may be more effective if targeted towards specific age groups. Water quality interventions may have limited effects on children under the age of six months, possibly because infants are experiencing limited exposure to household drinking water before weaning.[40] Interventions that reduce exposure from hand- and/or object-mouthing may be targeted at children aged 12-24 months, where we found the strongest associations between fomite rinses and diarrhea and linear growth, and hand rinses and diarrhea. Our log-stratified analyses indicated that diarrhea and linear growth burdens were driven by high levels of fecal contamination. It is important to note, that this does not necessarily suggest low levels of fecal contamination are unimportant for child health, since this finding may have resulted from misclassification of exposure due to limits of detections at low FIB concentrations. In other stratified analyses we found no differences by FIB used, in contrast to a prior systematic review that found a significant association between levels of *E. coli* in drinking water and diarrhea but not for fecal coliforms.[2] We found that overall effect estimates were driven by high levels of fecal contamination, but note that this finding could in part be induced by limits of detection at low levels of fecal contamination leading to misclassifications of exposure.

The limitations of this study reflect shortcomings in enteric exposure assessments in WaSH research. First, despite the vast number studies attempting to evaluate the impact of WaSH on health, few actually measure fecal contamination in the environment despite the fact that fecal exposure is a necessary intermediate step along the pathway to health effects. Those that do are largely confined to drinking water despite evidence suggesting that food and exposures along other pathways, such as soil, are likely contributors to overall enteric exposure.[41,42] Second, we were not able to obtain sufficient IPD for some of the pre-specified stratified analyses. We found a stronger effect from fecal contamination in drinking water on child diarrhea in urban areas, but most available datasets had limited IPD for urban settings, except for drinking water and diarrhea and a small amount of urban data for child hands and diarrhea. Understanding the relationship between fecal contamination and child health is a priority in rapidly urbanizing low-income countries. We also had limited data for young children under the age of six

months, resulting in either no estimates or estimates with high levels of uncertainty for this age group. Third, there is risk of bias from the outcome assessments because of reliance on caregiver reported diarrhea and anthropometric measurements. Reported diarrhea is subject to courtesy and recall bias,[6,43] and anthropometric measurements are also prone to measurement error, particularly when data collection staff are not blinded to interventions.[44] Fourth, studies generally use FIB data rather than actual pathogens as a proxy for health-related fecal contamination, an indicator with well documented shortcomings.[45] Fifth, none of the included studies estimated fecal contamination along all included pathways, so we were limited to modeling each pathway individually without adjusting for fecal contamination along the other pathways in our models. While these factors limit the inferences that can be drawn from the data, they also suggest the need for significant improvements in the manner in which WaSH research assesses the impact of interventions on fecal exposure.

Even though our findings provide evidence of associations between fecal contamination and adverse child health outcomes, we were not able to control for all possible factors that could affect the modeled exposure-outcome associations, so there is risk of uncontrolled confounding in our results. It is also important to note that we used measured household-level FIB concentrations and fly density as a proxy for individual-level exposure with no data on the extent to which children are actually exposed to specific disease-causing pathogens. These type of proxy measures are prone to introducing heterogeneity to exposure-outcome effect estimates as well as bias the associations, commonly towards the null.[46] So while we found associations between domestic fecal contamination and adverse child health outcomes, improved measures of exposure – including direct measures of known or suspected enteric pathogens – would better characterize exposure-outcome relationships, and could inform intervention design and evaluation through this characterization.

3.6 References

1 Gundry S, Wright J, Conroy R. A systematic review of the health outcomes related to household water quality in developing countries. *J Water Health* 2004; **2**: 1–13.

2 Gruber JS, Ercumen A, Colford JM. Coliform Bacteria as Indicators of Diarrheal Risk in Household Drinking Water: Systematic Review and Meta- Analysis. *PLoS One* 2014; **9**: e107429.

3 Hodge J, Chang HH, Boisson S, Collin SM, Peletz R, Clasen T. Assessing the association between thermotolerant coliforms in drinking water and diarrhea: An analysis of individual-level data from multiple studies. *Environ Health Perspect* 2016; **124**: 1560–7.

4 Pickering AJ, Ercumen A, Arnold BF, *et al.* Fecal Indicator Bacteria along Multiple Environmental Transmission Pathways (Water, Hands, Food, Soil, Flies) and Subsequent Child Diarrhea in Rural Bangladesh. *Environ Sci Technol* 2018; **52**: 7928–36.

5 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; **340**. DOI:10.1136/bmj.c221.

6 Arnold BF, Galiani S, Ram PK, *et al.* Practice of Epidemiology Optimal Recall Period for Caregiver-reported Illness in Risk Factor and Intervention Studies: A Multicountry Study. *Am J Epidemiol* 2013; **177**: 361–71.

7 Sclar GD, Penakalapati G, Amato HK, *et al.* Assessing the impact of sanitation on indicators of fecal exposure along principal transmission pathways: A systematic review. *Int J Hyg Environ Health* 2016; **219**: 709–23.

8 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**: 344–52.

9 Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 2017; **37**: 4302–15.

10 WHO. WHO Child Growth Standards: Length/Height-for-age, Weight-forage, Weight-for-length, Weight-for-height, and Body Mass Index-for age: Methods and Development. 1. 2006 DOI:10.4067/S0370-41062009000400012.

11 Peel MC, Finlayson BL, Mcmahon TA. Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci* 2007; **11**: 1633–44.

12 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–60.

13 US EPA. Guidance on Selecting Age Groups for Monitoring and Assessing Childhood Exposures to Environmental Contaminants. Washington, DC, 2005 DOI:EPA/630/P-03/003F.

14 R Core Team. R: A language and environment for statistical computing. 2019. https://www.r-project.org/.

15 Wells G, Shea B, O'Connell D, *et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2011. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed June 12, 2019).

16 Voss PH, Rehfuess EA. Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. *J Epidemiol Community Health* 2013; **67**: 98–104.

17 Arnold BF, Khush RS, Ramaswamy P, *et al.* Causal inference methods to study nonrandomized,

preexisting development interventions. *PNAS* 2010; **107**: 22605–10.

18 Benjamin-Chung J, Amin N, Ercumen A, *et al.* A Randomized Controlled Trial to Measure Spillover Effects of a Combined Water, Sanitation, and Handwashing Intervention in Rural Bangladesh. *Am J Epidemiol* 2018; **187**: 1733–44.

19 Ercumen A, Naser AM, Unicomb L, Arnold BF, Colford Jr. JM, Luby SP. Effects of Source- versus Household Contamination of Tubewell Water on Child Diarrhea in Rural Bangladesh: A Randomized Controlled Trial. *PLoS One* 2015; **10**: e0121907.

20 Kirby MA, Nagel CL, Rosa G, *et al.* Use, microbiological effectiveness and health impact of a household water filter intervention in rural Rwanda—A matched cohort study. *Int J Hyg Environ Health* 2017; **220**: 1020–9.

21 Kirby MA, Nagel CL, Rosa G, *et al.* Effects of a large-scale distribution of water filters and natural draft rocket-style cookstoves on diarrhea and acute respiratory infection: A cluster-randomized controlled trial in Western Province, Rwanda. *PLOS Med* 2019; **16**: e1002812.

22 Luby SP, Halder AK, Huda TM, *et al.* Microbiological Contamination of Drinking Water Associated with Subsequent Child Diarrhea. *Am J Trop Med Hyg* 2015; **93**: 904–11.

23 Patil SR, Arnold BF, Salvatore AL, *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**. DOI:10.1371/journal.pmed.1001709.

24 Peletz R, Simuyandi M, Sarenje K, *et al.* Drinking water quality, feeding practices, and diarrhea among children under 2 years of HIV-positive mothers in peri-urban Zambia. *Am J Trop Med Hyg* 2011; **85**: 318–26.

25 Peletz R, Simunyama M, Sarenje K, *et al.* Assessing Water Filtration and Safe Storage in Households with Young Children of HIV-Positive Mothers: A Randomized, Controlled Trial in Zambia. *PLoS One* 2012; **7**: e46548.

26 Pickering A, Swarthout J, Mureithi M, *et al.* Can individual and integrated water, sanitation, and handwashing interventions reduce fecal contamination in the household environment? Evidence from the WASH Benefits cluster-randomized trial in rural Kenya. *bioRxiv [Preprint]* 2019; : 731992.

27 Reese H, Routray P, Torondel B, *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; **0**: 1–11.

28 Sinharoy SS, Schmidt W-P, Wendt R, *et al.* Effect of community health clubs on child diarrhoea in western Rwanda: cluster-randomised controlled trial. *Lancet Glob Heal* 2017; **5**: e699–709.

29 Boisson S, Kiyombo M, Sthreshley L, Tumba S, Makambo J, Clasen T. Field assessment of a novel household-based water filtration device: A randomised, placebo-controlled trial in the democratic Republic of Congo. *PLoS One* 2010; **5**: 1–10.

30 Boisson S, Stevenson M, Shapiro L, *et al.* Effect of Household-Based Drinking Water Chlorination on Diarrhoea among Children under Five in Orissa, India: A Double-Blind Randomised Placebo-Controlled Trial. *PLoS Med* 2013; **10**: e1001497.

31 Brown J, Sobsey MD, Loomis D. Local Drinking Water Filters Reduce Diarrheal Disease in Cambodia: A Randomized, Controlled Trial of the Ceramic Water Purifier. *Am J Trop Med Hyg* 2008; **79**: 394–400.

32 Clasen T, Parra GG, Boisson S, Collin S. Household-based ceramic water filters for the prevention

of diarrhea: A randomized, controlled trial of a pilot program in Colombia. *Am J Trop Med Hyg* 2005; **73**: 790–5.

33 Clasen T, Boisson S, Routray P, *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* 2014; **2**: e645–53.

34 Ercumen A, Pickering AJ, Kwong LH, *et al.* Do Sanitation Improvements Reduce Fecal Contamination of Water, Hands, Food, Soil, and Flies? Evidence from a Cluster-Randomized Controlled Trial in Rural Bangladesh. *Environ Sci Technol* 2018; **52**: 12089–97.

35 Mattioli MC, Boehm AB, Davis J, Harris AR, Mrisho M, Pickering AJ. Enteric pathogens in stored drinking water and on caregiver's hands in Tanzanian households with and without reported cases of child diarrhea. *PLoS One* 2014; **9**: e84939.

36 Devamani C, Norman G, Schmidt W-P. A simple microbiological tool to evaluate the effect of environmental health interventions on hand contamination. *Int J Environ Res Public Health* 2014; **11**: 11846–59.

37 Vujcic J, Ram PK, Hussain F, *et al.* Toys and toilets: Cross-sectional study using children's toys to evaluate environmental faecal contamination in rural Bangladeshi households with different sanitation facilities and practices. *Trop Med Int Heal* 2014; **19**: 528–36.

38 WHO. Guidelines for Drinking-water Quality: 4th Edition. Geneva, 2017 http://apps.who.int/iris/bitstream/handle/10665/254637/9789241549950-eng.pdf?sequence=1 (accessed June 12, 2018).

39 Pickering AJ, Null C, Winch PJ, *et al.* The WASH Benefits and SHINE trials: interpretation of WASH intervention effects on linear growth and diarrhoea. *Lancet Glob Heal* 2019; **7**: e1139–46.

40 VanDerslice J, Popkin B, Briscoe J. Drinking-water quality, sanitation, and breast-feeding: their interactive effects on infant health. *Bull World Health Organ* 1994; **72**: 589–601.

41 Robb K, Null C, Teunis P, Yakubu H, Armah G, Moe CL. Assessment of Fecal Exposure Pathways in Low-Income Urban Neighborhoods in Accra, Ghana: Rationale, Design, Methods, and Key Findings of the SaniPath Study. *Am J Trop Med Hyg* 2017; **97**: 1020–32.

42 George C, Oldja L, Biswas S, *et al.* Geophagy is Associated with Environmental Enteropathy and Stunting in Children in Rural Bangladesh. *Am J Trop Med Hyg* 2015; **92**: 1117–24.

43 Schmidt W-P, Arnold BF, Boisson S, *et al.* Epidemiological methods in diarrhoea studies—an update. *Int J Epidemiol* 2011; **40**: 1678–92.

44 Hunter PR, Bartram J, Cairncross CS. Comment on "Randomized Intervention Study of Solar Disinfection of Drinking Water in the Prevention of Dysentery in Kenyan Children Aged under 5 Years". *Environ Sci Technol* 2009; **43**: 3035.

45 Leclerc H, Mossel DAA, Edberg SC, Struijk CB. Advances in the Bacteriology of the Coliform Group: Their Suitability as Markers of Microbial Water Safety. *Annu Rev Microbiol* 2001; **55**: 201–34.

46 Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010; **340**: c2289.

4 Chapter 4 – Research Aim 2: Exposure measurement error and the characterization of child exposure to fecal contamination in drinking water

Manuscript authors: Frederick G.B. Goddard, Howard H. Chang, Thomas Clasen, Jeremy A. Sarnat

4.1 Context

The second research aim of this dissertation built on the findings from the first aim by critically evaluating the evidence generated using exposure measurement error frameworks more commonly employed in other areas of environmental health, including air pollution epidemiology. We used the extensive water quality database we compiled for the first aim and sought to describe how random exposure measurement error, i.e. the difference that is not systematically biased in a particular direction between the observed exposure using proxies such as household-level fecal contamination and the true exposure, might affect estimated exposure-outcome effect estimates.

4.2 Background

Methods to characterize exposure to fecal contamination from contaminated drinking water in low- and middle-income settings have typically been limited to estimating fecal loading in the environment using indicators of fecal contamination.[1] These are proxy measures, in the sense that rather than measuring the actual ingestion of enteric pathogens associated with fecal contamination, they infer exposure by measuring fecal indicator bacteria (FIB) concentrations in the environment and are heavily based on assumptions on the interactions of individuals with that environment. In addition to the well-documented shortcomings of using FIB as a proxy for enteric pathogens,[2] the difference between the observed exposure, i.e. the exposure assigned from these proxy measures, and the true exposure represents a potential form of exposure measurement error. Other areas of environmental health have detailed how exposure measurement error may introduce bias and uncertainty in estimated exposure-outcome relationships, thereby obscuring true associations.[3] Correspondingly, fields such as air pollution

epidemiology have seen increased emphasis on understanding errors associated with differences between individual or personal exposures and other proxy exposure measures used in health effects modeling.[4]

The findings from the first research aim of this dissertation suggest that household-level FIB concentrations in drinking water, as a proxy measure of individual-level enteric pathogen ingestion from drinking water, are associated with both reported diarrhea (OR 1.09; 95%CI 1.03, 1.15) and lower height-for-age Z (HAZ) scores (HAZ -0.04; 95%CI -0.06, -0.01). Notably, we also observed moderate heterogeneity among studies in the strengths of association for both the diarrhea (I^2= 34%; 95% CI 0 – 62%) and growth analyses (I^2= 19%; 95% CI 0 – 63%). A primary limitation of this analysis was potential error in outcome measurement. In these studies, most of the data were collected for non-blinded intervention evaluations where caregiver reported diarrhea and linear growth measures are subject to participant and enumerator bias.[5–7] Another possible source of bias and uncertainty in our analysis may be in part due to errors in the assigned exposure. Prior research studying the effects of exposure measurement error from proxy measures of exposure in air pollution epidemiology suggest that these proxy measures can introduce uncertainty and bias risk estimates towards no observed effect.[8–10] Similar effects of exposure measurement error have also been shown in other areas, such as chemical exposures[11] and diastolic blood pressure measurements.[12] It is possible that findings from our IPD analysis may exhibit similar uncertainty and bias due to exposure measurement error from the use of FIB concentrations in household drinking water as a proxy for personal exposure to enteric pathogens.

One way of broadly categorizing types of exposure measurement error is to assess the distribution of error around the measurement of interest. These distributions are typically referred to as being either Berkson or classical, and may contain elements of both.[13] Briefly, classical measurement error occurs when the error is independent from *true* exposure and will generally originate from misclassification of individual-level exposure. In contrast, Berkson measurement error is independent from the *observed* exposure, or exposure indicator used in an epidemiologic model, and typically originates from assigning aggregate or mean exposure values in place of individual-level data. In general, classical error has been shown to lead

to increased uncertainty and an attenuation of the dose-response relationship, leading to biased effects towards a null hypothesis. This is also known as regression dilution bias. Berkson error typically does not introduce bias, but can lead to an increase in uncertainty and a subsequent loss in power to detect association where they truly exist. These types of measurement errors are random, in the sense that the difference between observed and true exposure averages out to zero and is not consistently biased in a particular direction.[14]

Theoretically, there are a number of different potential sources of error in fecal exposure assessments that could introduce Berksonian and/or classical-type errors. Some examples include 1) temporal or spatial variability in water quality; 2) assigning household- or community-level water quality to individuals; 3) the use of FIBs as proxies for enteric pathogens; 4) processing errors (i.e. during sample collection, transport or laboratory instrumentation errors). This is not an exclusive list of sources of error and each source could be further broken down in to underlying sources of error. Here, we introduce a conceptualized model of potential sources of measurement error using a formal measurement error framework.[15] We demonstrate how such a framework might be evaluated and its ability to quantify the relative contributions of measurement error, using empirical data of drinking water across several global low- and middle-income settings.

4.2.1 Exposure measurement error framework

Although quite limited, prior research on the effects of exposure measurement error on waterborne disease epidemiology has found preliminary evidence of regression dilution bias between 14 and 57% from the use of FIB on the relationship between fecal contamination in recreational water and swimming-associated illness.[16] Another study found attenuation between 35 and 45% from spatiotemporal variability in rainfall data in subsequent associations between heavy rainfall and diarrhea.[17] However, these studies focused on single components of error, and we did not find any formal discussions pertaining to multiple sources of exposure error for enteric pathogen exposure in the peer-reviewed literature. Here, we introduce an exposure measurement error framework to conceptualize multiple possible components of

error. We adapt a framework for fecal contamination exposure in drinking water based on an approach presented in Zeger et al. to distinguish sources of exposure measurement error.[15] While our analysis borrows from extensive work examining air pollution health effects, we intend that this research aim serves as an initial step for discussing ways of incorporating estimates of error for enteric exposures in a low-income setting context.

Zeger et al. consider sources of error in the assignment of ambient air quality to a population from central monitoring sites, a more distal measure of exposure than attempting to quantify personal exposures. While this framework is contextual to time series studies of air pollution health effects, here, we use this approach to identify and conceptualize sources of measurement error from drinking water fecal contamination exposure assessments. We see parallels in the proxy measures commonly used for fecal contamination exposure assessments to central-site air pollution exposure assessments, although we do not claim this framework includes all possible sources of error for water quality applications. In the current example involving fecal contamination exposure and response, as in the air pollution design settings that Zeger et al. used, technical and logistical constraints as well as limited resources lead to an inability to obtain measures of true personal exposure x for individual i at time t. Instead, exposure may be estimated by measuring household FIB concentrations z at time t. In Figure 4-1 we summarize the differences between measured FIB concentrations z_t, the only component in this framework that is actively measured, and true personal exposure to fecal contamination x_{it}, as a proxy for enteric pathogen exposure.

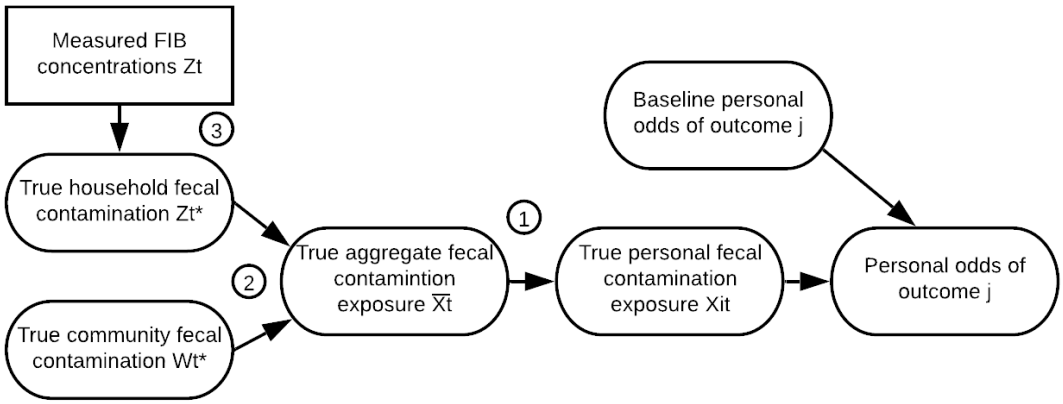


Figure 4-1: Fecal contamination in drinking water exposure measurement error framework (adapted from[15])

There is a difference between measured FIB concentrations in household drinking water z_t and true personal exposure to fecal contamination x_{it} – the exposure measurement error – which we split in to three components of error in accordance with the Zeger et al. framework (Equation 4-1).

$$x_{it} = z_t + (x_{it} - \bar{x}_t) + (\bar{x}_t - z_t^*) + (z_t^* - z_t) \qquad (4\text{-}1)$$

where (x_{it} - $\bar{x}_t$) (x_{it}-$\bar{x}_t$)describes error from the difference in aggregate fecal contamination exposure across a population $\bar{x}_t$ (i.e. members of a household) and personal exposure $x_{it}$$x_{it}$; ($(\bar{x}_t$-$z_t^*)$ $\bar{x}_t - z^*_t$) describes measurement error from assigning household water fecal contamination z_t^* as the exposure and not considering other exposures, such as exposure to fecal contamination experienced in the community w_t^* that may make up the aggregate exposure $\bar{x}_t$ $\bar{x}_t$across a population; and (z^*_t - z_t) describes measurement error from the difference in measured household water FIB concentrations z_t as an indicator of fecal contamination and the true levels of fecal contamination in household drinking water z_t^*.

Our current analysis sought to address the second and third components of this framework with the goal of examining how they may affect exposure-outcome relationships for exposure to fecal contamination in drinking water. This analysis does not seek to validate any single exposure characterization method, but rather to describe potential sources of error in current methods to help inform future methods development. In conducting this analysis, we sought to assess evidence and magnitude of exposure measurement error from: 1) from assigning household-level FIB concentrations z_t^* as the exposure and

not considering community-level FIB concentrations w_t^* (component 2) in the FIB-diarrhea relationship; and 2) from using single FIB measures z_t $z_t$compared to repeated longitudinal measures z_t^* (component 3) in the FIB-linear growth relationship.

4.3 Methods

4.3.1 Data

We used data from 19 studies conducted in South America, Sub-Saharan Africa, South and South-East Asia.[18–36] We requested permission from data owners for use of these data for this study. Eligible datasets included variables describing FIB concentrations in household drinking water, child age and intervention status. We included children aged 0-72 months. Datasets also included unique identifiers for each community, household and child. For the diarrhea analysis, we defined community water fecal contamination levels for a given household on a specific day as the median household water FIB concentrations of all other households in its community on the same day. We generated a variable for each city, or collection of communities in rural areas, describing whether water quality data was collected in a wet or dry season month using the same methodology as for Research Aim 12atching single household water FIB concentration observations z_t to child survey data if they were collected on the same day or up to seven days before the survey was conducted, and then generated different scenarios for the aggregate fecal contamination $\bar{x}_t$ by incorporating the median community water fecal contamination. For the growth analysis we used household water FIB concentrations data in wide format by matching all available water samples collected over the course of a child's life up to the day anthropometric measurements were taken.

4.3.2 Analytical approach

We used a two-tiered analytical approach to examine evidence and magnitude of random exposure measurement error. First, in a simulated analysis we randomly assigned health outcomes (diarrhea cases and HAZ scores) to each observation with an estimated exposure and then regressed those outcomes on the error-prone exposure variables, represented by the measured proxies of exposure. Exposure was

assigned based on household water FIB concentration measurements, so the simulations retained existing correlations between communities, households and individuals. Evidence and magnitude of exposure measurement error was assessed by estimating the attenuation factor associated with the error-prone exposure variable.[37] Second, we evaluated findings from our simulations by using empirical health outcome data from the same datasets and regressing it on both the estimated exposure and error-prone exposure variables. All analyses were conducted in R version 3.6.[38]

Household versus Community Exposure

Phase I: We simulated the effect of exclusively assigning household water fecal contamination z_t^* for individual exposure, if estimated exposure is actually a combination of both household and community water fecal contamination w_t^*w$_t^*$, by:

1. Randomly generating diarrhea cases for each included child with a combination of household- and community-level drinking water fecal contamination as the aggregate drinking water fecal contamination $\bar{x}_t$ experienced by a child, using the Bernoulli distribution where the log odds of diarrhea d_{ijkl} for child i in household j in community k and study l is given by:

$$logit(d_{ijkl}) = \beta_0 + \beta_1 FIB_{ij} + \beta_2 Treat_{ij} + \beta_3 Age_i + \beta_4 Season_{ijk} + \mu_{ijkl} + \mu_{ijk}$$

$$\mu_{ijkl} \sim N(0, 0.6);\ \mu_{ijk} \sim N(0, 0.3)$$

$$p_{ijkl} = \frac{e^{d_{ijkl}}}{1 + e^{d_{ijkl}}}$$

$$d_{ijkl} = bernoulli(ijkl, p_{ijkl})$$

2. Assuming that 1) community-level drinking water fecal contamination is represented by the median household water FIB concentrations in all other community households; 2) baseline odds of diarrhea β_0 $\beta_0$for this population is 0.15; 3) odds of diarrhea for 1-log$_{10}$ higher FIB concentrations in drinking water β_1 $\beta_1$is 1.5; and 4) odds of diarrhea for children receiving an intervention $\beta_2\beta_2$, child age β_3 (in years) and for data collected in the wet compared to the dry season $\beta_4\beta_4$ are 0.9, 0.8 and 1.2, respectively. Effect estimates were broadly based on model

outputs from our IPD analyses, although we assumed higher odds of diarrhea for FIB concentrations in drinking water because we hypothesize that the effect estimate for the exposure-outcome relationship in our IPD analysis may have been suffering from regression dilution bias[14]. The model accounted for clustering at the study-level μ_{ijkl} and community-level $\mu_{ijk}.\mu_{ijk}$.

3. Fitting multilevel generalized mixed effects models with the assigned diarrheal cases and replacing the combined household and community drinking water FIB concentrations $\bar{x}_t$ with household-level FIB concentrations z_t^* exclusively as the error-prone exposure variable.
4. Calculating the attenuation associated with the estimated log odds of diarrhea (β_1^*) from assigning household-level FIB concentrations z_t^* exclusively as the error-prone exposure variable, compared to the assigned log odds of diarrhea ($\beta_1 = \log(1.5)$) if combined household and community drinking water FIB concentrations $\bar{x}_t \bar{x}_t$ represent the exposure: $\alpha = \frac{\beta_1^*}{\beta_1} = \frac{\beta_1^*}{\log(1.5)}$
5. Repeating simulations for a range of exposure scenarios by adding community water fecal contamination w_t^* in 10% increments, starting with 100% household water fecal contamination and ending with 100% community water fecal contamination representing the estimated exposure.
6. Stratifying the combined analysis: As reported under Research Aim 1, estimated odds of diarrhea for a 1-$\log_{10}$ increase in FIB concentrations in drinking water was higher in urban compared to rural settings, so we stratified the simulation by urban versus rural areas to differentiate whether exclusively assigning household drinking water fecal contamination z_t^* may introduce more error in one setting compared to the other.
7. Conducting sensitivity analyses: Assessed the effects our assumptions had on the simulation findings by repeating the simulations with a) higher and lower assumed odds of diarrhea for higher FIB concentrations in drinking water and; b) using the highest and lowest community water FIB concentrations instead of the median.

Phase II: To evaluate findings from the simulations we applied empirical diarrhea data, by:

1. Beginning with household water fecal contamination z_t^* as the estimated exposure and fitting a multilevel generalized mixed effects model to estimate the odds of diarrhea for 1-$\log_{10}$ higher FIB concentrations in household drinking water.
2. Replacing household water fecal contamination z_t^* with community water fecal contamination w_t^* in 10% increments and fitting the same regression model with each new exposure assignment.
3. Calculating the attenuation associated with the log odds of diarrhea (β_1^*) from assigning household-level FIB concentrations exclusively as the error-prone exposure variable, compared to effect estimates that combine household/community water fecal contamination (β_1).
4. Stratifying the analysis: In addition to stratifying by rural versus urban areas, we also stratified by children aged 0-23 and 24-72 months, to consider how child mobility may modify the effect of assigning community water quality to exposure. We hypothesized that children aged 0-23 months are mostly non-ambulatory and spend the majority of their time within the confines of their home, and pre-school children aged 24-72 months are ambulatory and spend their time both in their home and within the confines of the community.

4.3.3 Single versus Multiple Samples

Phase I: We simulated the effect of assigning a single measure of FIB concentrations in drinking water z_t as the error-prone exposure variable, by

1. Randomly generating expected HAZ scores with the estimated household water fecal contamination z_t^* represented by the median household water FIB concentrations from repeat samples, using the using the following model where the difference in HAZ-scores HAZ_{ijkl} for child i in household j in community k and study l is given by:

$$HAZ_{ijkl} = \beta_0 + \beta_1 FIB_{ij} + \beta_2 Treat_{ij} + \beta_3 Age_i + \mu_{ijkl} + \in_{ijkl};$$

$$\mu_{ijkl} \sim N(0, 0.5);\ \in_{ijkl} \sim N(0, 1)$$

2. Assuming that the 1) mean baseline HAZ score β_0 in this population is -1.6; 2) difference in HAZ score for 1-$\log_{10}$ higher median FIB concentrations $\beta_1\beta_1$ is -0.2; 3) difference in HAZ score for children receiving an intervention $\beta_2\beta_2$ and for child age $\beta_3\beta_3$ (years) are 0.1 and -0.05; and 4) HAZ scores follow a normal distribution. The model accounted for clustering at the study-level $\mu_{ijkl}\mu_{ijkl}$.
3. Fitting multilevel generalized mixed effects models with the assigned HAZ scores and replacing the estimated household water fecal contamination z_t^* represented by the median household water FIB concentrations from repeat samples with a randomly chosen single measure of household water FIB concentrations z_t $z_t$as the error-prone exposure variable.
4. Calculating the attenuation associated with the estimated difference in HAZ score (β_1^*) from randomly choosing a single measure of water quality z_t as the error-prone exposure variable, compared to the assigned difference in HAZ score (β_1 = -0.2), if the estimated exposure is represented by repeat samples of household water fecal contamination z_t^* $\alpha = \frac{\beta_1 *}{\beta_1} = \frac{\beta_1 *}{-0.2}$
5. Repeating the simulations for children with at least two, three or four matched household water FIB concentration measures making up the median household water fecal contamination z_t^*. We did not have sufficient data to conduct these simulations with more than four matched water samples.
6. Stratifying the analysis: Previous research has found that fecal contamination in drinking water sources in low-income countries is higher in the wet season compared to the dry season,[39] so we stratified these simulations by season to examine whether error introduced from variability in water quality is greater in wet season months compared to dry season months.
7. Conducting sensitivity analyses: Assessed the effects our assumptions had on the simulation findings by repeating the simulations with a) higher and lower assumed difference in HAZ scores

for higher FIB concentrations in drinking water and; b) using the highest and lowest drinking water FIB concentrations from the repeat samples instead of the median.

Phase II: To evaluate findings from the simulations we applied empirical linear growth data for a subset of children in our dataset where HAZ scores were available, by:

1. Beginning by fitting multilevel generalized mixed effects models with the median household water fecal contamination z_t^* from repeated measures of household FIB concentrations as the exposure variable. Repeating this for two, three and four repeat measures.
2. Fitting the same models after randomly selecting a single measure of household water FIB concentrations z_t as the error-prone exposure variable from the repeat measures.
3. Calculating the attenuation associated with the estimated difference in HAZ score (β_1^*) from randomly selecting a single measure of household water FIB concentrations z_t as the error-prone exposure variable, compared to effect estimate (β_1) from the median of repeat samples of household water fecal contamination z_t^* as the exposure variable.
4. Stratifying the analysis: We conducted the same stratification by season as we did for the simulations.

4.4 Results

The dataset we compiled included studies of varying sizes with FIB data for drinking water available from 98 to 2,137 households per study (Table 4-1). Four studies included only cross-sectional water sample collection, but most had collected repeated water samples over time with samples typically being collected monthly, quarterly or annually. To evaluate findings from our simulations, matched diarrhea data were available from all included studies and matched growth data were available from seven studies.

Table 4-1: Summary Characteristics of included studies

Study ID	Country	Nr. of households sampled	Samples per household	Matched diarrhea data	Matched growth data
Arnold, 2010	India	139	1 – 3	X	X
Benjamin-Chung, 2018	Bangladesh	959	1	X	
Boisson, 2010	DRC	102	1	X	
Boisson, 2013	India	2,100	1 – 4	X	
Brown, 2008	Cambodia	177	1 – 11	X	
Clasen, 2005	Colombia	98	1 – 5	X	
Clasen, 2014	India	2,137	1 – 8	X	X
Davis, in prep.	Tanzania	1,207	1 – 6	X	
Ercumen 2015	Bangladesh	907	1 – 3	X	
Kirby, 2017	Rwanda	265	1 – 2	X	
Kirby, Nagel, 2019	Rwanda	1,568	1 – 3	X	
Luby, 2015	Bangladesh	810	1 – 7	X	
Patil, 2015	India	783	1	X	X
Peletz, 2011	Zambia	232	1	X	
Peletz, 2012	Zambia	116	1 – 12	X	
Pickering, Ercumen, 2018	Bangladesh	1,807	1 – 3	X	X
Pickering, 2019	Kenya	1,986	1 – 2	X	X
Reese, 2019	India	765	1 – 4	X	X
Sinharoy, 2017	Rwanda	1,744	1 – 2	X	X

Household versus Community Exposure

The simulations comparing household to community exposure included 37,119 observations (82% rural, 18% urban) from 16 of the included studies, with between one and 12 observations per child. Studies by Arnold et al., Brown et al. and Pickering, Ercumen et al. were excluded from these simulations because the data were not compatible for estimating median community water FIB concentrations (i.e. specific sample collection dates were not available or only one sample was collected in every community). The simulations found that if children experienced exposure to fecal contamination in drinking water outside of their household, then using household FIB concentrations z_t^* exclusively as the error-prone exposure variable may lead to an attenuation in the observed FIB-diarrhea relationship even at low levels of community exposure (Figure 4-2). If the aggregate exposure $\overline{x}_t$ was represented by 90% household and 10% community exposure, we found that the estimated odds of diarrhea for 1-$\log_{10}$ higher FIB

concentrations in drinking water, using household FIB concentrations in drinking water $z_t^* z_t^*$ as the error-prone exposure variable, were OR = 1.32 compared to the assigned odds of diarrhea OR = 1.50 (α= 0.69). This trend continued as the assumed aggregate exposure $\bar{x}_t$ consisted of increasing levels of community exposure w_t^*. If the aggregate exposure was represented by 100% community exposure, we found that the estimated odds of diarrhea, using household water as the error-prone exposure variable, were OR = 1.06 (α= 0.15). Our findings were consistent between urban and rural areas. Findings from the sensitivity analyses are provided in Appendix D (Tables D-1 and D-2).

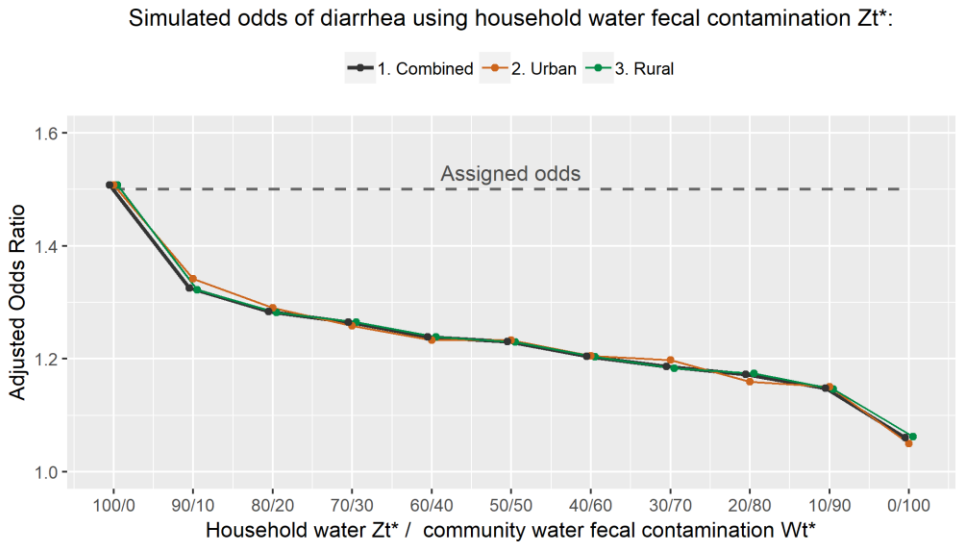


Figure 4-2: Estimated odds of diarrhea for 1-log_{10} higher FIB concentrations in drinking water when only household water FIB concentrations z_t^* was used as the exposure variable for different household-community exposure scenarios and the corresponding attenuation factors. Exposure scenarios begin with 100% of exposure assumed to be within the household and subsequently replacing household exposure with community exposure w_t^* in 10% increments.

For the evaluation with empirical diarrhea data we used the same dataset as we did for the simulations, with the exception of using field reported diarrhea instead of assigning diarrhea cases. In the combined analysis, we found that household water fecal contamination z_t^* assigned exclusively as the error-prone exposure variable attenuated the association between FIB concentrations in drinking water and diarrhea in comparison to a mixture of household and community water fecal contamination up to assigning 20% household and 80% community water FIB concentrations (Figure 4-3). However, this attenuation was not

as pronounced as suggested in the simulations, with the greatest attenuation factor found to be α= 0.79 (60% household, 40% community exposure). The stratified results suggested that in urban areas the attenuation from using household water fecal contamination z_t^* was limited to scenarios up to 70% household and 30% community water quality, and in rural areas it extended to 10% household and 90% community water fecal contamination .

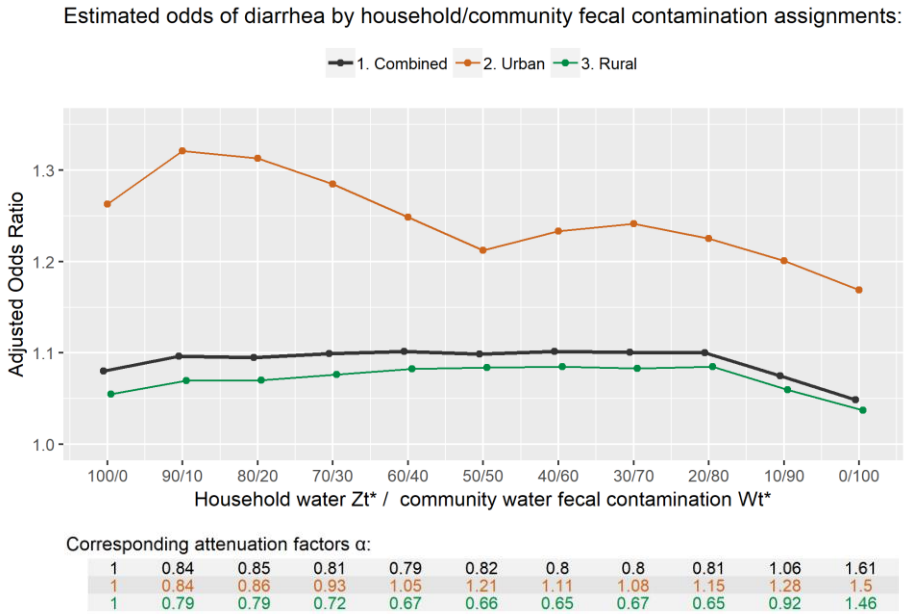


Figure 4-3: Estimated odds of diarrhea for 1-log_{10} higher FIB concentrations in drinking water for different household-community exposure assignments and the corresponding attenuation factors for estimates derived from 100% household exposure compared to the community assignments.

After stratifying by age, the odds of diarrhea for 1-log_{10} higher FIB concentrations in drinking water for children aged 0-23 months were similar to the odds of diarrhea for children aged 24-72 months when household water fecal contamination z_t^* was assigned as the exposure variable. However, when replacing household water fecal contamination with community water fecal contamination w_t^* w_t^* in 10% increments, there was a trend of higher odds of diarrhea for children aged 24-72 months but not for children aged 0-23 months for up to 20% household and 80% community water exposure (Figure 4-4).

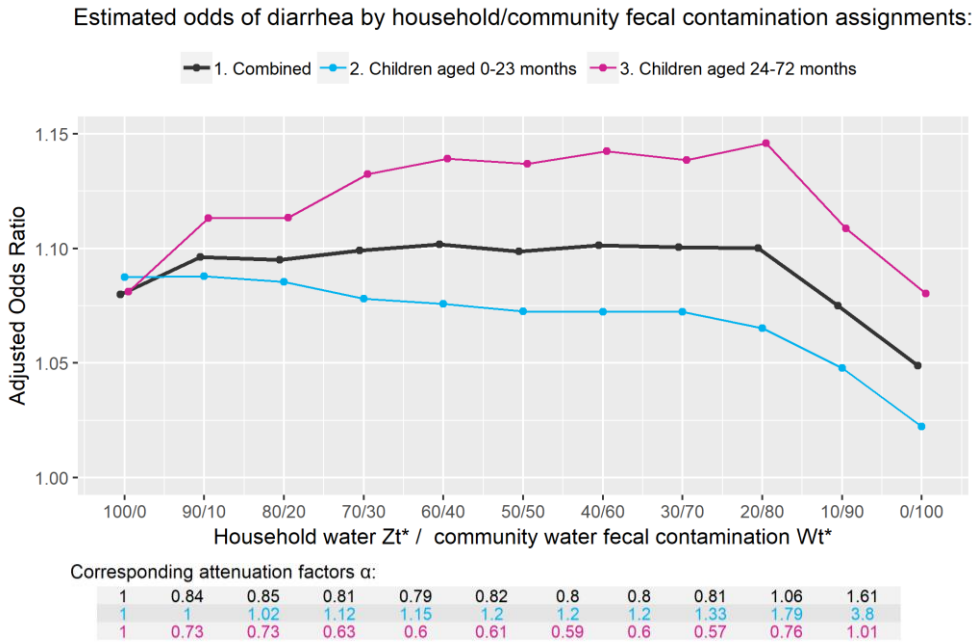
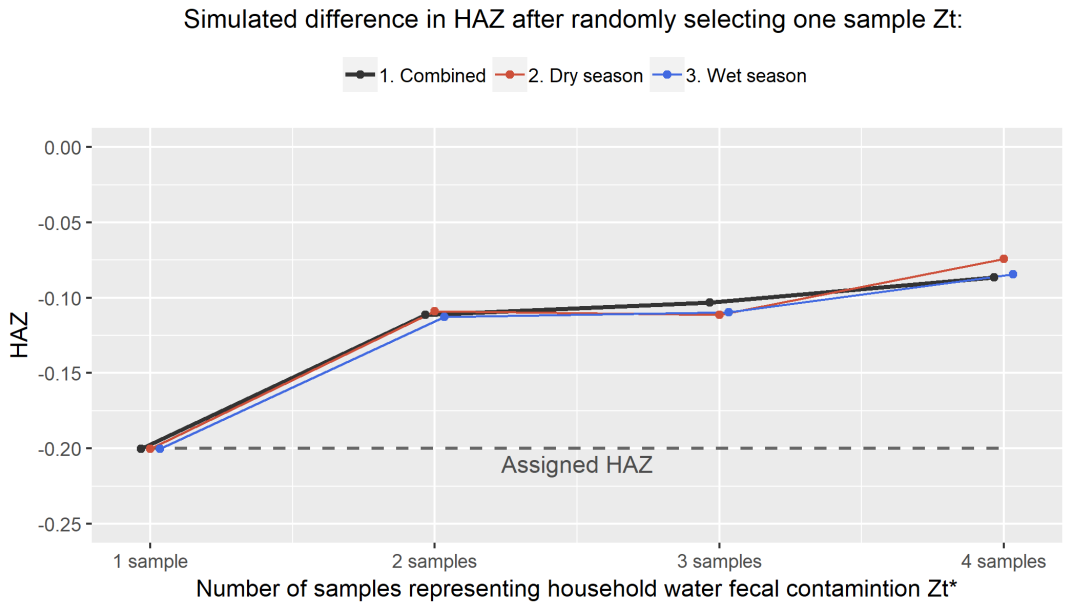


Figure 4-4: Estimated odds of diarrhea for 1-log$_{10}$ higher FIB concentrations in drinking water for different household-community exposure assignments, stratified by non-ambulatory and ambulatory pre-school aged children, and the corresponding attenuation factors for estimates derived from 100% household exposure compared to the community assignments.

Single versus Multiple Samples

The simulations comparing the effects of defining exposure with a single water sample compared to multiple samples included 24,806 unique children from the 19 included studies that had one or more matched FIB estimates in drinking water. These simulations indicated that using FIB concentrations from single water samples z_t $z_t$compared to the median of multiple samples attenuated estimated differences in HAZ scores associated with fecal contamination in drinking water, with similar findings between wet and dry season months (Figure 4-5). If the household water fecal contamination z_t^* was represented by the median of two samples, then randomly selecting one of the two samples almost halved the observed difference in HAZ scores associated with FIB concentrations in drinking water from the assigned HAZ = -0.20 to HAZ = -0.11 (α= 0.56). This finding was more pronounced when household water fecal contamination z_t^* was represented by three (HAZ = -0.10; α= 0.52) or four samples (HAZ = -0.09; α= 0.43). Findings from the sensitivity analyses are detailed in Appendix D (Tables D-3 and D-4).



Corresponding attenuation factors α:

1	0.56	0.52	0.43
1	0.55	0.56	0.37
1	0.56	0.55	0.42

Figure 4-5: Estimated difference in HAZ scores for 1-log_{10} higher FIB concentrations in drinking water when only a single water sample z_t z_t was used as the exposure variable for four scenarios if 1, 2 3 or 4 samples represented the household water fecal contamination z_t^* and the corresponding attenuation factors.

For the evaluation of these simulations we were limited by empirical linear growth data availability, with linear growth data available for 3311, 743 and 233 children with two, three and four or more matched water samples, respectively. As a result, we were not able to stratify this analysis by season as we did with the simulations, and the baseline effect estimates and corresponding uncertainties around these estimates vary by group because they represent different samples (Figure 4-6). The difference in HAZ scores associated with higher FIB concentrations in drinking water was consistently closer to zero (i.e., no effect) when using a single sample z_t $z_t$compared to the median of multiple samples. Similar to findings from the simulations, using a single sample compared to the median of two samples approximately halved the estimated difference in HAZ scores associated with FIB concentrations in drinking water ($\alpha = 0.56$), and this was more pronounced for the median of three or four samples ($\alpha= 0.54$; $\alpha = 0.38$).

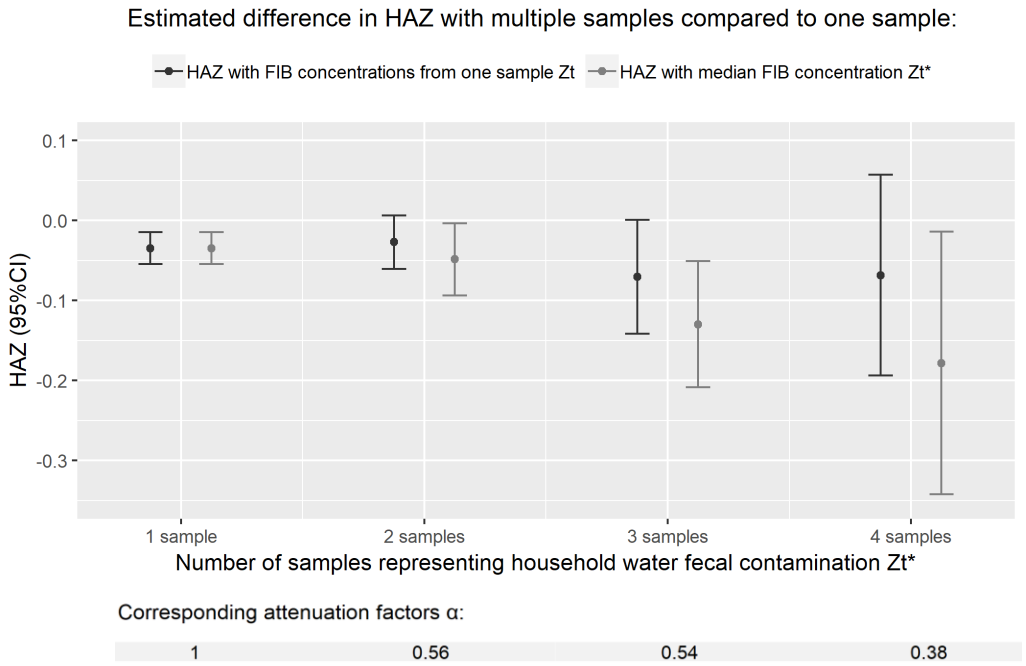


Figure 4-6: Estimated difference in HAZ scores for 1-log_{10} higher FIB concentrations in drinking water when only a single household water fecal contamination z_t was used as the exposure variable if 1, 2 3 or 4 samples were available, and the corresponding attenuation factors.

4.5 Discussion

We adapted and introduced a framework to assess measurement error when characterizing child exposure to fecal contamination in drinking water, i.e. the difference between exposure assigned by proxy measures of exposure and the true exposure experienced by an individual. These frameworks can help prioritize current research gaps by identifying areas within fecal exposure assessments that are limited or missing, and by quantifying components of error that are most critical to biases in waterborne disease epidemiology. Ideally, generating improved exposure data can lead to a better understanding of the true associations between fecal contamination along different pathways and child health. This analysis primarily serves as an initial effort to apply an exposure measurement error framework within the field of enteric exposure science. In so doing, we aspire to understand the presence and magnitude of several sources of measurement error. Our analyses showed how components of error may attenuate estimated exposure-outcome relationships using empirical data from an extensive dataset of studies collected in low- and middle-income settings. Our findings provide indication that the previously reported odds of

diarrhea and reduction in HAZ scores associated with fecal contamination in drinking water reported in Research Aim 1, may be prone to regression dilution bias and thus may be underestimating true exposure-outcome relationships.

We introduced three different components of exposure measurement error. The first component may emerge from assigning household water fecal contamination data to individual household members who interact with their environment differently. Substantial heterogeneity of between-child interactions with their domestic environment has been shown in both urban and rural settings for different age groups in the 0-5 year age range.[40,41] Additionally, differential drinking water ingestion rates by age can lead to heterogeneity in the ingested doses of fecal contamination,[42] and infants may experience very limited exposure to household water from ingestion before weaning.[43] This may lead to a differences in dose-response between members of the same household. To test how the first component of this exposure measurement error framework can be applied, small controlled panel studies are needed to generate estimates of personal exposure and compare those to household-level estimates.[8]

The second component of error may occur when exposure to fecal contamination in drinking water outside of the household is not incorporated in to exposure assessments. A recent study characterizing fecal exposure in Accra, Ghana as part of the SaniPath research program, reported widespread fecal contamination in both domestic and public domains.[44] Measurement error from assigning household water fecal contamination as the exposure does not only depend on the presence of fecal contamination in the public domain but the study population's interaction with water in that domain. To our knowledge, no published studies have quantified child exposure to contaminated water in different microenvironments in the domestic and public domains, but time-activity analyses in air pollution studies have long been conducted for exposure assessments,[45] and have shown that children spend extensive amounts of time outside of their domestic environment.[46]

Findings from our diarrhea simulations suggest that if children are experiencing exposure to fecal contamination in drinking water outside of their households, then using household water FIB

concentrations as a proxy for their overall exposure may result in attenuated FIB-diarrhea effect estimates. While evaluating these results, we found that this attenuation may be more pronounced in children above the age of two. This suggests that children under the age of two may be experiencing most of their exposure within the confines of their homes, so household-level exposure assessments may be appropriate for this age group. However, for older ambulatory children exposure outside of the home might be more readily considered.

The third component of error may emerge from limited precision associated with methods to characterize FIB concentrations,[47,48] i.e. from variability in water quality measurements due to sampling and laboratory processing methods, but can also stem from temporal differences in water quality. FIB levels in household water can vary on a weekly, daily and even hourly basis.[49] For an outcome such as diarrhea that is normally acute, the biologically-relevant household water fecal contamination levels might be representative of the fecal contamination levels during the incubation period of enteric pathogens found in water, which depending on the pathogen can vary from a matter of hours to up to a month.[50] If water samples are collected on the same day as diarrheal disease data, the measured FIB concentrations on that day may not be representative of the biologically-relevant fecal contamination in the lead up to a diarrhea episode. These discrepancies could be due to environmental factors, such as short term weather changes like extreme rainfall events,[51] or human factors, such as water treatment behavior change in response to a diarrhea episode.[21,52] For chronic outcomes such as child growth, the biologically-relevant household water fecal contamination likely needs to consider longer term fecal contamination exposure , which may not be adequately represented by single or a few repeat measurements of household water fecal contamination, due to short term and seasonal variability in fecal contamination in drinking water.[53]

Our simulations suggest that long-term household water fecal contamination may not be adequately represented by a single sample and hence can result in attenuations of the FIB-growth relationship. These results were consistent with our evaluations using empirical growth data. While our previously reported IPD analysis found a significant association between fecal contamination and child linear growth, 70% of

the sample population only had a single matched water quality measure available to characterize exposure. The findings from this analysis imply that the reported effect sizes may be attenuated, and fecal exposure assessments may consider characterizing fecal contamination using multiple longitudinally samples to estimate more biologically-relevant exposure.

The results from our analyses need to be interpreted with caution. First, due to data availability we were limited to applying this framework to two sources of error. There are many more possible sources of error that we were not able to consider here, such as assigning household-level exposures to individuals and the use of FIB as proxies for enteric pathogens. Second, this analysis was limited to quantifying the effects of measurement error on the magnitudes of health effects and not on the precision of those effect estimates. Uncertainty in health effect estimates introduced by exposure measurement error may obscure associations where they exist, thus increasing the likelihood of false-negative findings in the exposure-outcome relationship. Third, the current framework is limited to drinking water, but there are a number of other important fecal-oral transmission pathways, such as hands, food, soil, fomites and flies. Findings from our IPD analysis suggest that fecal contamination along select pathways is associated with child diarrhea and growth, so a similar framework could be applied to other pathways to test whether those findings may have suffered from regression dilution bias. Fourth, we did not have access to repeated water samples within the shorter timeframe of pathogen incubation periods for acute gastroenteritis, so were not able to quantify measurement error in the FIB-diarrhea relationship from the use of single samples used to estimate household water fecal contamination.

Our results suggest that exposure measurement error can contribute to attenuated fecal exposure-outcome relationships for outcomes that are typically acute, such as diarrhea, as well as for more chronic outcomes such as linear growth. Fecal exposure assessments in drinking water may consider exposure outside of the household as well as attempting to characterize fecal contamination with repeat samples to account for variability in water quality. They may leverage measurement error frameworks to design exposure

assessments that are more proximal to the true exposure experienced by individuals, which in turn may inform the design of more effective interventions to reduce waterborne disease burdens.

4.6 References

1 Sclar GD, Penakalapati G, Amato HK, *et al.* Assessing the impact of sanitation on indicators of fecal exposure along principal transmission pathways: A systematic review. *Int J Hyg Environ Health* 2016; **219**: 709–23.

2 Leclerc H, Mossel DAA, Edberg SC, Struijk CB. Advances in the Bacteriology of the Coliform Group: Their Suitability as Markers of Microbial Water Safety. *Annu Rev Microbiol* 2001; **55**: 201–34.

3 Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998; **55**: 651–6.

4 Sarnat JA, Wilson WE, Strand M, Brook J, Wyzga R, Lumley T. Panel discussion review: session 1 — exposure assessment and related errors in air pollution epidemiologic studies. *J Expo Sci Environ Epidemiol* 2007; **17**: S75–82.

5 Hunter PR, Bartram J, Cairncross CS. Comment on "Randomized Intervention Study of Solar Disinfection of Drinking Water in the Prevention of Dysentery in Kenyan Children Aged under 5 Years". *Environ Sci Technol* 2009; **43**: 3035.

6 Schmidt W-P, Arnold BF, Boisson S, *et al.* Epidemiological methods in diarrhoea studies—an update. *Int J Epidemiol* 2011; **40**: 1678–92.

7 Arnold BF, Galiani S, Ram PK, *et al.* Practice of Epidemiology Optimal Recall Period for Caregiver-reported Illness in Risk Factor and Intervention Studies: A Multicountry Study. *Am J Epidemiol* 2013; **177**: 361–71.

8 Schwartz J, Sarnat JA, Coull BA, Wilson WE. Effects of exposure measurement error on particle matter epidemiology: a simulation using data from a panel study in Baltimore, MD. *J Expo Sci Environ Epidemiol* 2007; **17**: S2–10.

9 Goldman GT, Mulholland JA, Russell AG, *et al.* Impact of exposure measurement error in air pollution epidemiology: effect of error type in time-series studies. *Environ Heal* 2011; **10**: 61.

10 Pennington AF, Strickland MJ, Klein M, *et al.* Measurement error in mobile source air pollution exposure estimates due to residential mobility during pregnancy. *J Expo Sci Environ Epidemiol* 2017; **27**: 513–20.

11 Perrier F, Giorgis-Allemand L, Slama R, Philippat C. Within-subject Pooling of Biological Samples to Reduce Exposure Misclassification in Biomarker-based Studies. *Epidemiology* 2016; **27**: 378–88.

12 MacMahon S, Peto R, Cutler J, *et al.* Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; **335**: 765–74.

13 Weisskopf MG, Webster TF. Trade-offs of Personal Versus More Proxy Exposure Measures in Environmental Epidemiology. *Epidemiology* 2017; **28**: 635–43.

14 Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BM* 2010; **340**: c2289.

15 Zeger SL, Thomas D, Dominici F, *et al.* Exposure Measurement Error in Time-Series Studies of Air Pollution: Concepts and Consequences. *Environ Health Perspect* 2000; **108**: 419–26.

16 Fleisher JM. The Effects of Measurement Error on Previously Reported Mathematical Relationships between Indicator Organism Density and Swimming-Associated Illness: A

Quantitative Estimate of the Resulting Bias. *Int J Epidemiol* 1990; **19**: 1100–6.

17 Levy MC, Collender PA, Carlton EJ, *et al.* Spatiotemporal Error in Rainfall Data: Consequences for Epidemiologic Analysis of Waterborne Diseases. *Am J Epidemiol* 2019; **188**: 950–9.

18 Arnold BF, Khush RS, Ramaswamy P, *et al.* Causal inference methods to study nonrandomized, preexisting development interventions. *PNAS* 2010; **107**: 22605–10.

19 Benjamin-Chung J, Amin N, Ercumen A, *et al.* A Randomized Controlled Trial to Measure Spillover Effects of a Combined Water, Sanitation, and Handwashing Intervention in Rural Bangladesh. *Am J Epidemiol* 2018; **187**: 1733–44.

20 Kirby MA, Nagel CL, Rosa G, *et al.* Effects of a large-scale distribution of water filters and natural draft rocket-style cookstoves on diarrhea and acute respiratory infection: A cluster-randomized controlled trial in Western Province, Rwanda. *PLOS Med* 2019; **16**: e1002812.

21 Luby SP, Halder AK, Huda TM, *et al.* Microbiological Contamination of Drinking Water Associated with Subsequent Child Diarrhea. *Am J Trop Med Hyg* 2015; **93**: 904–11.

22 Patil SR, Arnold BF, Salvatore AL, *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**. DOI:10.1371/journal.pmed.1001709.

23 Peletz R, Simuyandi M, Sarenje K, *et al.* Drinking water quality, feeding practices, and diarrhea among children under 2 years of HIV-positive mothers in peri-urban Zambia. *Am J Trop Med Hyg* 2011; **85**: 318–26.

24 Peletz R, Simunyama M, Sarenje K, *et al.* Assessing Water Filtration and Safe Storage in Households with Young Children of HIV-Positive Mothers: A Randomized, Controlled Trial in Zambia. *PLoS One* 2012; **7**: e46548.

25 Pickering AJ, Ercumen A, Arnold BF, *et al.* Fecal Indicator Bacteria along Multiple Environmental Transmission Pathways (Water, Hands, Food, Soil, Flies) and Subsequent Child Diarrhea in Rural Bangladesh. *Environ Sci Technol* 2018; **52**: 7928–36.

26 Pickering A, Swarthout J, Mureithi M, *et al.* Can individual and integrated water, sanitation, and handwashing interventions reduce fecal contamination in the household environment? Evidence from the WASH Benefits cluster-randomized trial in rural Kenya. *bioRxiv [Preprint]* 2019; : 731992.

27 Reese H, Routray P, Torondel B, *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; **0**: 1–11.

28 Sinharoy SS, Schmidt W-P, Wendt R, *et al.* Effect of community health clubs on child diarrhoea in western Rwanda: cluster-randomised controlled trial. *Lancet Glob Heal* 2017; **5**: e699–709.

29 Boisson S, Kiyombo M, Sthreshley L, Tumba S, Makambo J, Clasen T. Field assessment of a novel household-based water filtration device: A randomised, placebo-controlled trial in the democratic Republic of Congo. *PLoS One* 2010; **5**: 1–10.

30 Boisson S, Stevenson M, Shapiro L, *et al.* Effect of Household-Based Drinking Water Chlorination on Diarrhoea among Children under Five in Orissa, India: A Double-Blind Randomised Placebo-Controlled Trial. *PLoS Med* 2013; **10**: e1001497.

31 Brown J, Sobsey MD, Loomis D. Local Drinking Water Filters Reduce Diarrheal Disease in Cambodia: A Randomized, Controlled Trial of the Ceramic Water Purifier. *Am J Trop Med Hyg* 2008; **79**: 394–400.

32 Clasen T, Parra GG, Boisson S, Collin S. Household-based ceramic water filters for the prevention of diarrhea: A randomized, controlled trial of a pilot program in Colombia. *Am J Trop Med Hyg* 2005; **73**: 790–5.

33 Clasen T, Boisson S, Routray P, *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* 2014; **2**: e645–53.

34 Mattioli MC, Boehm AB, Davis J, Harris AR, Mrisho M, Pickering AJ. Enteric pathogens in stored drinking water and on caregiver's hands in Tanzanian households with and without reported cases of child diarrhea. *PLoS One* 2014; **9**: e84939.

35 Ercumen A, Naser AM, Unicomb L, Arnold BF, Colford Jr. JM, Luby SP. Effects of Source-versus Household Contamination of Tubewell Water on Child Diarrhea in Rural Bangladesh: A Randomized Controlled Trial. *PLoS One* 2015; **10**: e0121907.

36 Kirby MA, Nagel CL, Rosa G, *et al.* Use, microbiological effectiveness and health impact of a household water filter intervention in rural Rwanda—A matched cohort study. *Int J Hyg Environ Health* 2017; **220**: 1020–9.

37 Fuller WA. Measurement Error Models. New York: John Wiley & Sons, 1987 https://onlinelibrary-wiley-com.proxy.library.emory.edu/doi/pdf/10.1002/9780470316665 (accessed July 22, 2019).

38 R Core Team. R: A language and environment for statistical computing. 2019. https://www.r-project.org/.

39 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.

40 Teunis PFM, Reese HE, Null C, Yakubu H, Moe CL. Quantifying Contact with the Environment: Behaviors of Young Children in Accra, Ghana. *Am J Trop Med Hyg* 2016; **94**: 920–31.

41 Kwong LH, Ercumen A, Pickering AJ, Unicomb L, Davis J, Luby SP. Age-related changes to environmental exposure: variation in the frequency that young children place hands and objects in their mouths. *J Expo Sci Environ Epidemiol* 2019; : 1.

42 US EPA. Exposure Factors Handbook: 2011 Edition. 2011 DOI:EPA/600/R-090/052F.

43 VanDerslice J, Popkin B, Briscoe J. Drinking-water quality, sanitation, and breast-feeding: their interactive effects on infant health. *Bull World Health Organ* 1994; **72**: 589–601.

44 Robb K, Null C, Teunis P, Yakubu H, Armah G, Moe CL. Assessment of Fecal Exposure Pathways in Low-Income Urban Neighborhoods in Accra, Ghana: Rationale, Design, Methods, and Key Findings of the SaniPath Study. *Am J Trop Med Hyg* 2017; **97**: 1020–32.

45 Branco PTBS, Alvim-Ferraz MCM, Martins FG, Sousa SI V. The microenvironmental modelling approach to assess children's exposure to air pollution – A review. *Environ Res* 2014; **135**: 317–32.

46 Devakumar D, Semple S, Osrin D, *et al.* Biomass fuel use and the exposure of children to particulate air pollution in southern Nepal. *Environ Int* 2014; **66**: 79–87.

47 Harmel RD, Hathaway JM, Wagner KL, *et al.* Uncertainty in monitoring E. coli concentrations in streams and stormwater runoff. *J Hydrol* 2016; **534**: 524–33.

48 Gronewold AD, Wolpert RL. Modeling the relationship between most probable number (MPN) and colony-forming unit (CFU) estimates of fecal coliform concentration. *Water Res* 2008; **42**: 3327–34.

49 Levy K, Hubbard AE, Nelson KL, Eisenberg JNS. Drivers of Water Quality Variability in Northern Coastal Ecuador. *Environ Sci Technol* 2009; **43**: 1788–97.

50 European Centre for Disease Prevention and Control. Systematic review on the incubation and infectiousness/ shedding period of communicable diseases in children. Stockholm, 2016 DOI:10.2900/337026.

51 Guzman Herrador BR, de Blasio BF, MacDonald E, *et al.* Analytical studies assessing the association between extreme precipitation or temperature and drinking water-related waterborne infections: a review. *Environ Heal* 2015; **14**: 29.

52 Ercumen A, Arnold BF, Naser AM, Unicomb L, Colford JM, Luby SP. Potential sources of bias in the use of Escherichia coli to measure waterborne diarrhoea risk in low-income settings. *Trop Med Int Heal* 2016; **22**: 2–11.

53 Wilkes G, Edge T, Gannon V, *et al.* Seasonal relationships among indicator bacteria, pathogenic bacteria, Cryptosporidium oocysts, Giardia cysts, and hydrological indices for surface waters within an agricultural landscape. *Water Resarch* 2009; **43**: 2209–23.

5 Chapter 5 – Research Aim 3: Child salivary SIgA and its relationship to enteric infections and EED biomarkers in Maputo, Mozambique

Manuscript authors: Frederick G.B. Goddard, Jackie Knee, Trent Sumner, Rassul Nalá, Thomas Clasen, Joe Brown

5.1 Context

Research Aims 1 and 2 of this dissertation have focused on proxy methods of external enteric exposure, i.e. exposure that is assessed before enteric pathogens cross the human boundary, and their biological relevance to health outcomes of interest. This chapter goes beyond the human boundary to pilot methods to estimate internal exposure, i.e. characterizing exposure after ingestion of enteric pathogens. Specifically, we measured acute antibody responses in saliva, a biological matrix that may represent a promising alternative for infectious disease surveillance, and investigated associations with enteric infections and biomarkers of environmental enteric dysfunction (EED) found in matched stool samples.

5.2 Background

In low- and middle-income countries, children experience high prevalence of enteric infections[1,2] and subsequently suffer disproportionally from diarrheal disease morbidity.[3] Enteric infections are also associated with chronic gastrointestinal health outcomes, such as EED induced from repeated infections.[4] EED is a subclinical disorder associated with intestinal inflammation and a reduced ability to absorb nutrients.[5] Findings from a recent systematic review suggest that intestinal inflammation is linked to child stunting,[6] the burden of which is greatest in low- and middle-income countries.[7]

Characterizing enteric infections often relies on the collection of stool samples to use culture-based or molecular methods to detect enteric pathogen shedding,[8] or immunoassays to measure immunological responses to specific pathogens and concentrations of EED biomarkers.[9] Collecting, transporting and storing stool specimens can be resource intensive, so the use of stool can be especially challenging for epidemiological studies in resource-constrained settings.[10] Serum or saliva as alternative biological

matrices in immunological assays may present a promising alternative.[11] Serum has the disadvantages that sample collection requires trained individuals and its invasive procedure may lead to low response rates, especially among children,[12] and there is a non-zero risk of blood-borne pathogen transmission from respondents to sample collectors.[13] In contrast, saliva collection is non-invasive and requires only minimal training,[14] facilitating sample collection in large study populations of young children.[15]

One class of biomarkers that can be measured in saliva and may be of interest for gastrointestinal health are salivary antibodies. There are different antibody isotypes with varying functions in the response to gastrointestinal infections. Broadly, immunoglobulin A (IgA) and IgM are produced in response to acute infections, whereas IgG is typically produced later and can be indicative of chronic or historical infections.[16] Saliva has a very low concentration of IgM and IgG antibodies in comparison to IgA.[17] IgA is produced by plasma cells and released in secretory fluids, such as saliva, as secretory IgA (SIgA).[18] The amount of IgA that humans release is linked to the development of the systemic immune system, which evolves over the course of human life, beginning as an immature immune system during infancy that matures during childhood.[19] Accordingly, previous research on SIgA concentrations in saliva with age suggest a rapid increase during infancy,[20] followed by more gradual increases during early childhood, and a stabilization at adult levels during adolescence.[21]

The biological mechanisms underlying SIgA-mediated immunity are convoluted, with three separate immunological mechanisms active in protecting the intestinal epithelium from enteric infections. The first prevents pathogens from attaching to the intestinal epithelial cell barrier, most commonly by a process called immune exclusion.[22] Humans release three grams of SIgA in to the intestinal lumen every day,[23] where it acts as the first line of defense to protect the intestinal epithelium against pathogenic organisms.[24] SIgA is transported across the intestinal epithelium and into the lumen by binding to the polymeric immunoglobulin receptor (pIgR), where it binds to the pathogen and neutralizes its ability to attach to the intestinal epithelium and cause infection.[25] The second is specifically targeted at viruses, by neutralizing viruses inside the epithelial cell and preventing assembly/disassembly and exit from the epithelial cell.[25]

Third, if damage to the epithelial cell barrier has allowed pathogen invasion in to the lamina propria, SIgA can bind to the pathogen and transport it back across the epithelial boundary using the pIgR for excretion.[26]

Under the umbrella of the Maputo Sanitation trial (MapSan) in Maputo Mozambique, this chapter sought to estimate the differences in total salivary acute antibody responses among children experiencing concurrent enteric infections and concentrations of EED biomarkers detected in matched stool samples. We also compared salivary pathogen-specific acute antibody responses to the detection of specific pathogens in matched stool samples. Specifically, we quantified total non-specific SIgA in saliva to estimate differences in acute systemic immune responses in young children experiencing varying numbers of concurrent enteric infections, as well as different concentrations of biomarkers of local gut inflammation and permeability. We also compared anti-*Campylobacter jejuni* (*C. jejuni)* and anti-*Giardia* salivary SIgA concentrations in children with and without matched *Giardia*- or *Campylobacter*-positive stool samples.

5.3 Methods

5.3.1 Study setting and participants

MapSan was a controlled, before-and after trial of an urban sanitation intervention to reduce enteric infections and improve other health metrics in children in Maputo, Mozambique.[27] The study was located in informal settlements and enrolled children aged one to 48 months during the baseline phase between February 2015 and February 2016 and subsequently conducted 12- and 24-month follow-up surveys. Enrollment in MapSan was progressive, and all eligible, consenting children were enrolled during each survey phase (children aged one to 60 months at 12- or 24-month follow-up). MapSan enrolled 993 children at baseline, 320 at 12-month follow-up, and 304 at 24-month follow-up.[28] For this sub-study, we analyzed one saliva sample per child from a subset of 244 children collected at either 12- or 24-month follow-up.

Saliva samples assayed for total non-specific salivary SIgA were selected based on sufficient sample volume (>10μL), child age, availability of a matched stool sample, and the number of infections detected in that stool sample. We excluded samples from children under the age of 12 months due to the presence of maternal antibodies and lack of crevicular fluid (saliva excreted between the teeth and gums enriched with Ig).[29] Saliva samples were eligible if they were collected within ten days of matched stool samples. Due to the high prevalence of enteric infection in the MapSan cohort,[28] we selected all available saliva samples from children where no infections were detected in matched stool samples, if they also met the other eligibility criteria. A subset of the samples assayed for total SIgA were selected for the anti-*C. jejuni* and anti-*Giardia* SIgA assays, by selecting a balance of 'case' samples, i.e. samples that were positive for *Campylobacter or Giardia*, and 'control' samples i.e. samples that were negative for *Campylobacter* or *Giardia*.

5.3.2 Procedures

Saliva samples were collected during three cross-sectional household survey visits at the baseline, 12- and 24-month assessments of the MapSan study. Baseline samples were collected with cotton swabs and were subsequently not eligible for this study due to insufficient sample volume. During the 12- and 24-month assessments saliva samples were collected by rubbing Oracol saliva swabs (Malvern Medical Developments, Worcester, United Kingdom), along the child's gum for one minute to collect crevicular fluid. Samples were transported in a cooler and frozen at -80 °C until processed. To prepare samples for processing, we centrifuged saliva swabs at 2,000rcf for 10 minutes, before removing saliva from the sample collection tubes and recording sample volume. We excluded saliva samples visibly contaminated with serum.

We used enzyme-linked immunosorbent assays (ELISA) to process samples for total non-specific SIgA, anti-*C. jejuni* and anti-*Giardia* SIgA. Total SIgA was assayed in accordance with the manufacturer's directions (Salimetrics, Carlsbad, CA, USA). For anti-*Giardia* and anti-*C. Jejuni* SIgA we used commercially available ELISA kits validated for use with serum (IBL, Minneapolis, MN, USA and

Epitiope Diagnostics Inc., San Diego, CA, USA) and adapted these kits for use with saliva. Due to lower concentrations of SIgA in saliva compared to serum, we adapted the ELISA kits by changing sample dilutions. We processed samples for anti-*C. jejuni* SIgA at dilutions ranging from 1:2 to 1:6, and diluted samples assayed for anti-*Giardia* SIgA from 1:2 to 1:10, depending on sample availability. We made no other adaptations to the ELISA kits and tested anti-*C. jejuni* and anti-*Giardia* SIgA in accordance with the manufacturer's directions. We processed 70% of samples in replicate for total SIgA, and 100% in replicate for anti-*C. jejuni and* anti-*Giardia*. For replicate samples we excluded results where individual replicates were not within 20% of the replicate mean.

Laboratory methods to detect enteric pathogens and EED biomarkers in matched stool samples are reported elsewhere.[30] Briefly, stool samples collected for the MapSan study were analyzed for detection of 14 enteric pathogens using the molecular-based Luminex Gastrointestinal Pathogen Panels (GPP). The GPP included bacterial pathogens (*Campylobacter (C. jejuni, C. coli, and C. lari)* , *Clostridium difficile (C. difficile), Escherichia coli (E. coli)* O157, Enterotoxigenic *E. coli* (ETEC), Shiga-like toxin producing *E. coli* (STEC), *Shigella, Vibrio cholerae* (*V. cholerae)* and *Yersinia enterocolitica (Y. enterocolitica)*), protozoan pathogens (*Giardia*, *Cryptosporidium* and *Entamoeba histolytica (E. histolytica)*) and viral pathogens (Adenovirus 40/41, Norovirus GI/GII, and Rotavirus A). The same stool samples were analyzed for biomarkers of gut inflammation using ELISA assays. Biomarkers included myeloperoxidase and fecal calprotectin, both markers of neutrophil activity,[31] alpha-1 antitrypsin, a protein released during inflammation and marker of gut permeability,[32] and neopterin, a marker of T helper cell derived immune activation.[33]

5.3.3 Statistical analysis

We matched total SIgA measured in each saliva sample to individual data on enteric pathogens and EED biomarkers detected in stool samples collected within ten days of saliva samples. Most stool and saliva samples were collected on the same day or within 24 hours of each other, and all were modeled as cross-sectional matched samples. To account for non-normality of the data, we log-transformed all SIgA and

EED biomarker data. For our primary analyses, we used multilevel linear models to account for potential confounders and model cross-sectional associations between 1) total non-specific salivary SIgA concentrations and the number of concurrent enteric infections experienced by a child, 2) total non-specific salivary SIgA and concentrations of EED biomarkers, and 3) anti-*C. Jejuni* or anti-*Giardia* SIgA and presence of those pathogens in matched stool samples:

$$log_{10}(SIgA) = \beta_1 \inf(nr) + \beta_2 age + \beta_3 vol + \beta_4 rain \quad (5\text{-}1)$$

$$log_{10}(SIgA) = \beta_1 log_{10}(EED_b) + \beta_2 age + \beta_3 vol + \beta_4 rain + \beta_5 \inf(nr) \quad (5\text{-}2)$$

$$log_{10}(SIgA_p) = \beta_1 \mathrm{p} + \beta_2 age + \beta_3 vol + \beta_4 rain + \beta_5 \inf(nr) + \beta_6 log_{10}(SIgA) \quad (5\text{-}3)$$

Evidence in the literature suggests that salivary SIgA levels are affected by age,[34] salivary flow rate[35] and seasonality,[36] so we controlled for child age (in months), sample volume (in µL) and seasonality in all of our models. We controlled for seasonality by splitting cumulative rainfall during the 30 days before saliva sample collection in to terciles, where the first tercile (least rain) represents the reference level in our models. Rainfall data were obtained from the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (https://www.ncdc.noaa.gov/cdo-web/datatools/findstation). We also controlled for the number of concurrent enteric infections in our EED analyses, and for the number of concurrent infections and total non-specific SIgA in our pathogen-specific SIgA analyses. EED biomarkers, indexed by *b* in the model, were modeled individually, and so were the pathogen-specific SIgA results for each pathogen *p* (*Giardia and C. jejuni*).

For our secondary analyses, we did sensitivity analyses for the first two models to estimate the effects outliers had on our findings by excluding observations (for both total SIgA and EED biomarkers) that were 1.5 interquartile ranges below the lower quartile or above the upper quartile. We also conducted threshold analyses with the pathogen-specific SIgA data to detect potential positive anti-*C. jejuni* or anti-*Giardia* immune responses when *Campylobacter* or *Giardia* were detected in matched stool samples, where observations above the threshold suggest a positive immune response to the specific pathogen. We

defined an immunopositive threshold as the mean anti-*C. jejuni* or anti-*Giardia* SIgA concentration of the control samples used in the respective assays plus three standard deviations.[37]

5.3.4 Ethics

Field data collection staff obtained written informed consent from the parent or guardian of each study participant. The study protocol was approved by the Comité Nacional de Bioética para a Saúde (CNBS), Ministério da Saúde (333/CNBS/14), the Ethics Committee of the London School of Hygiene and Tropical Medicine (reference #8345), and the Institutional Review Board of the Georgia Institute of Technology (protocol #H15160). The MapSan study is registered at ClinicalTrials.gov (NCT02362932).

5.4 Results

5.4.1 Summary characteristics

We extracted 244 saliva samples with 216 samples presenting with sufficient sample volume and no visible blood to be eligible for testing (Table 5-1). Most of our saliva samples (89%) were collected within one day of stool sample collection. Child age ranged from 1 to 6.7 years with a median age of 2.5 years, with most of our data coming from children aged 1-2 years (63%) and less data for children aged 3-6 years (37%). Two samples were excluded from our analyses due to replicate rejection, but otherwise we found acceptable coefficients of variation between replicate samples. We found median total salivary SIgA levels of 54 μg/ml (IQR: 34, 85 μg/ml) in this study population, and total salivary SIgA was similar between children of different ages (Figure 5-1).

Table 5-1: Summary Characteristics.

	Total SIgA	Anti-*C. jejuni* SIgA	Anti-*Giardia* SIgA
Number of saliva samples:			
Extracted	244	-	-
Excluded due to insufficient volume	13	-	-
Excluded due to visible serum	15	-	-
Excluded due to replicate rejection	2	-	-
Included in analysis	214	66	64
Male child (%)	50	45	50
Child age in years – Median (IQR)	2.5 (1.8, 3.7)	2.5 (2.0, 3.5)	2.5 (1.8, 3.6)
Difference in days between saliva and stool sample collection – Median (IQR)	0 (-1, 1)	1 (-1, 1)	1 (0, 1)
Sample volume available in µL – Median (IQR)	175 (100, 300)	200 (150-300)	188 (100-300)
Total salivary SIgA levels in µg/ml – Median (IQR)	54 (34, 85)	55 (38, 84)	52 (28, 87)
Coefficient of variation between duplicate samples (%)	6.4	5.3	7.3

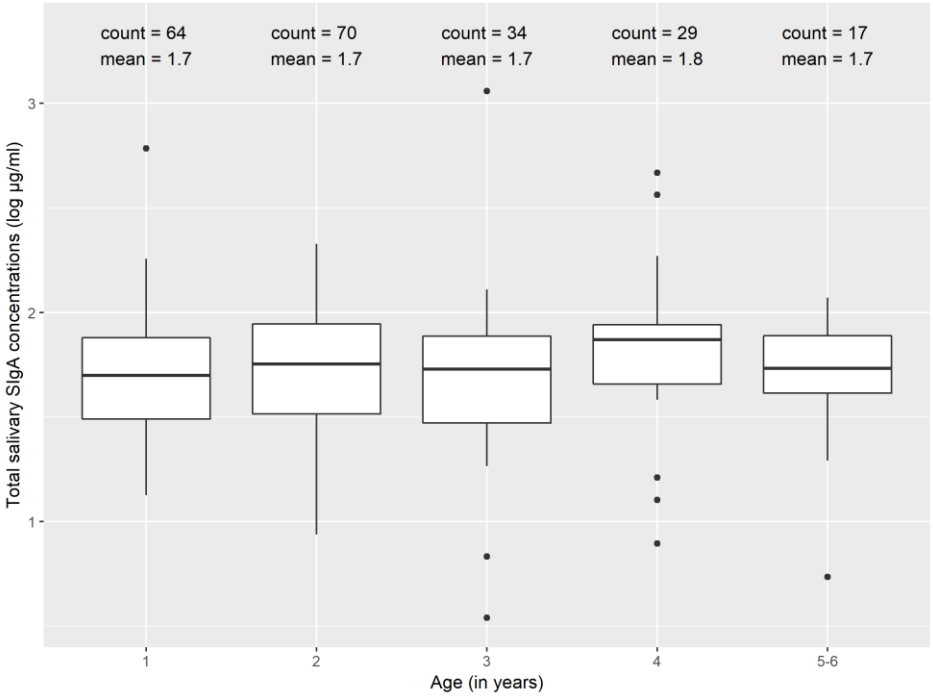


Figure 5-1: Total non-specific salivary SIgA concentrations (log µg/ml) by age

5.4.2 Total SIgA and enteric infections

This was a non-random sample, so the distribution of infections with specific pathogens for this sub-sample was not representative of the distribution found in the MapSan cohort (Figure 5-2). We did not see

a trend in higher total salivary SIgA from children experiencing none, one, two, three, or four to five concurrent infections detected in matched stool samples (Figure 5-3).

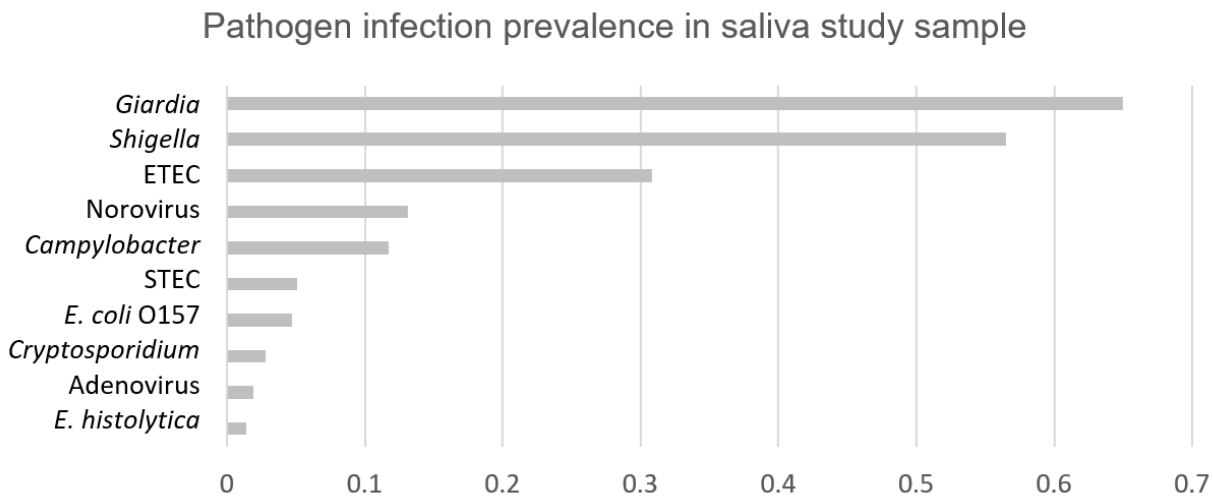


Figure 5-2: Prevalence of specific pathogens detected in stool samples in the MapSan saliva study sub-sample

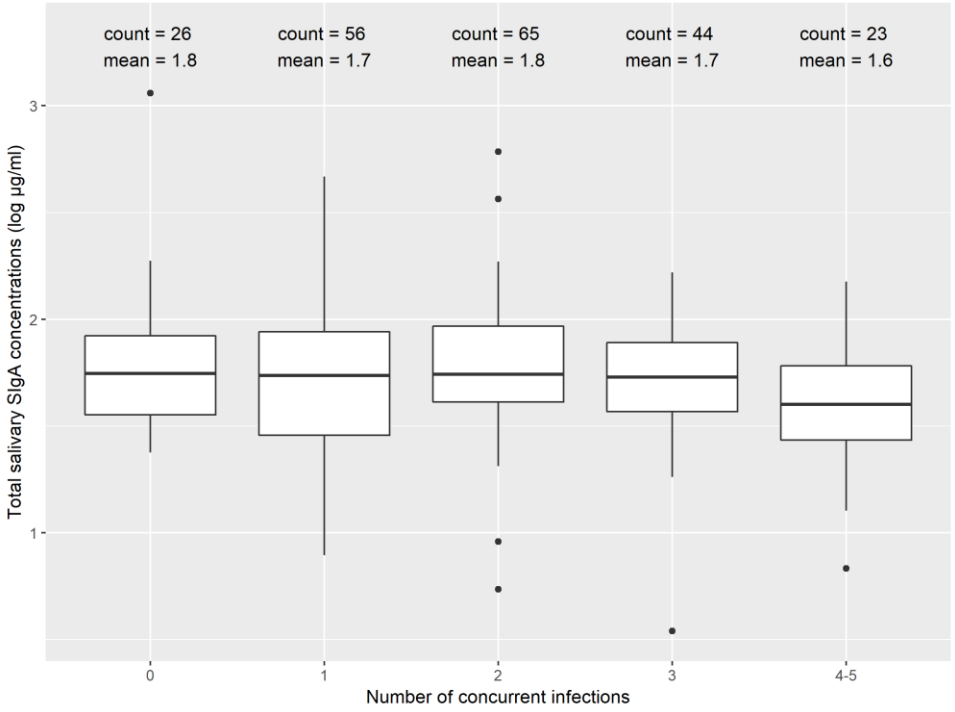


Figure 5-3: Total salivary SIgA concentrations stratified by children experiencing different numbers of concurrent infections.

Results from our statistical analysis suggested lower total salivary SIgA -0.04 log µg/ml (95% CI: -0.08 to -0.005 log µg/ml) for higher numbers of concurrent infections, although this association was weaker after

removing outliers (Table 5-2). Sample volume was also significantly negatively associated with total salivary SIgA, whereas we found no statistical difference of total salivary SIgA with child age or higher cumulative rainfall in the same model.

Table 5-2: Difference in total non-specific salivary SIgA with a higher number of concurrent infections, after controlling for age (in months), sample volume (in µL) and 30-day rainfall (in terciles).

	All samples (N = 214)			After removing outliers (N = 206)		
	Difference in total SIgA (log µg/ml)	**95% CI**	**p-value**	**Difference in total SIgA (log µg/ml)**	**95% CI**	**p-value**
Number of infections	-0.04	(-0.08, -5×10^{-3})	0.03	-0.03	(-0.06, 2×10^{-3})	0.07
Age (in months)	4×10^{-4}	(-2×10^{-3}, 3×10^{-3})	0.79	1×10^{-3}	(-1×10^{-3}, 3×10^{-3})	0.31
Sample volume (in µL)	-1×10^{-3}	(-9×10^{-4}, -3×10^{-4})	<0.001	-6×10^{-4}	(-8×10^{-4}, -3×10^{-4})	<0.001
Rainfall (terciles)	0.03	(-0.02, 0.08)	0.29	0.04	(-4×10^{-3}, 0.08)	0.07

5.4.3 Total SIgA and EED biomarkers

We found no association between total salivary SIgA and EED biomarkers of inflammation and permeability, in models including all samples and after removing outliers (Table 5-3).

Table 5-3: Difference in salivary SIgA for a unit difference in EED biomarkers found in stool, after controlling for age, sample volume, 30-day rainfall, and number of concurrent infections.

	All samples				After removing outliers			
EED Biomarker	**N**	**Difference in total SIgA (log µg/ml)**	**95% CI**	**p-value**	**N**	**Difference in total SIgA (log µg/ml)**	**95% CI**	**p-value**
Neopterin (log nmol/l)	188	0.02	(-0.09, 0.13)	0.75	180	-0.02	(-0.12, 0.07)	0.61
Myeloperoxidase (log ng/ml)	213	0.02	(-0.07, 0.12)	0.64	201	0.04	(-0.05, 0.12)	0.39
Calprotectin (log ng/ml)	211	0.02	(-0.06, 0.10)	0.68	202	4×10^{-3}	(-0.07, 0.07)	0.91
Alpha-1 antitrypsin (log ng/ml)	207	-0.08	(-0.17, 4×10^{-3})	0.06	196	-0.02	(-0.1, 0.06)	0.62

5.4.4 Pathogen-specific SIgA

The results for anti-*Giardia* and anti-*C. jejuni* SIgA are presented in arbitrary units, because the ELISAs we used established values of assay calibrators in arbitrary units (U/mL) due to a lack of gold standard concentrations for these assays. We found higher anti-*Giardia* SIgA in children with matched *Giardia*-positive stool samples compared to those with matched *Giardia*-negative stool samples 0.32 log U/ml (95% CI: 0.04 to 0.6 log U/ml) (Table 5-4). We found weaker evidence of higher anti-*C. jejuni* SIgA in children with matched *Campylobacter*-positive stool samples compared to those with matched *Campylobacter*-negative stool samples 0.27 log U/ml (95% CI: -0.04 to 0.57 log U/ml).

Table 5-4: Difference in anti-C.jejuni and anti-Giardia salivary SIgA in children with Giardia- or C. jejuni-positive matched stool samples compared to children with negative Giardia or C. jejuni stool samples, after controlling for age (in months), sample volume (in µL), 30-day rainfall (in terciles), number of concurrent infections and total SIgA (in log µg/ml) .

	C. Jejuni			Giardia		
	Difference in anti-C. Jejuni SIgA (log U/ml)	**95% CI**	**p-value**	**Difference in anti-Giardia SIgA (log U/ml)**	**95% CI**	**p-value**
Positive GPP result	0.27	(-0.04, 0.57)	0.09	0.32	(0.04, 0.6)	0.03
Age (in months)	$8x10^{-3}$	($-1x10^{-3}$, 0.02)	0.10	$-3x10^{-3}$	(-0.01, $6x10^{-3}$)	0.52
Sample volume (per unit increase in µL)	$-7x10^{-5}$	($-1x10^{-3}$; $1x10^{-3}$)	0.90	$1x10^{-3}$	($-4x10^{-5}$, $2x10^{-3}$)	0.06
Rainfall (terciles)	-0.12	(-0.29, 0.05)	0.16	0.08	(-0.08, 0.23)	0.33
Number of infections	-0.08	(-0.23, 0.08)	0.34	0.09	(-0.11, 0.29)	0.38
Total SIgA (per unit increase in log µg/ml)	0.59	(0.16, 1.0)	0.01	1.6	(1.1, 2.0)	<0.001

In our threshold analyses, we found one positive immune response from the anti-*C. jejuni* SIgA assay among children with matched *Campylobacter*-positive stool samples, and one positive immune response among children with *Campylobacter*-negative stool samples (Figure 5-4). For the anti-*Giardia* SIgA assay we found two positive immune responses among children with matched Giardia-positive stool samples, and one positive immune response among children with *Giardia*-negative stool samples.

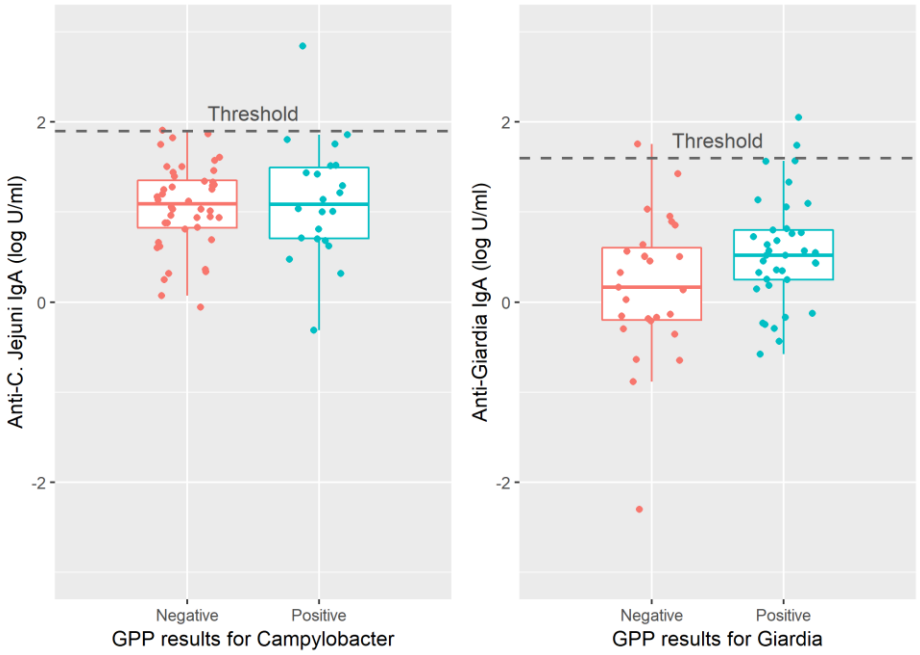


Figure 5-4: Anti-C. jejuni and anti-Giardia SIgA concentrations (log U/ml) stratified by children with negative and positive matched Campylobacter or Giardia stool samples.

5.5 Discussion

This study measured total non-specific salivary SIgA and tested its association with enteric infections and biomarkers of gastrointestinal inflammation and permeability found in matched stool samples from children aged one to six years in informal settlements in Maputo, Mozambique. We found no differences in total salivary SIgA concentration between children of different ages and lower total SIgA with higher numbers of concurrent infections detected in matched stool samples. We found no relationship between total SIgA and EED biomarkers, suggesting that systemic immune responses in this population were not associated with local gut inflammation or permeability.

The negative association between total SIgA levels in saliva and the number of concurrent infections experienced by a child, detected in matched stool samples, was unexpected. Since IgA is secreted in response to pathogenic organisms, we expected to find a positive relationship between SIgA levels and the number of concurrent infections. One possible hypothesis that could explain why we found the contrary, is that the number of concurrent infections with enteric pathogens may contribute to child

malnutrition[38,39] and severe malnutrition is associated with lower levels of SIgA in children.[40] So in this study population, high prevalence of infection and concurrent infections may be linked to a suppression of the immune system. It is notable that in addition to finding a negative relationship between total SIgA and the number of concurrent infections, we also found no difference in total SIgA concentrations between children of different ages. Age (in months) was not a significant confounder in any of our models, and we also did not see a trend of higher median total salivary SIgA concentrations after stratifying on age (in years). This is in contrast to previous findings from other studies. Evidence from Estonia, Sweden, Iceland and Israel suggests that total salivary SIgA levels are three to four times higher in children aged five to six years compared to one-year old children.[21,41,42] Taken together, these findings suggest that immunological development in this study sample of children living in an urban slum in Maputo, Mozambique, experiencing high rates of enteric infections may be inhibited during early life stages.

We are limited in the conclusions we can draw from our anti-*C. jejuni* and anti-*Giardia* SIgA findings. We found associations between anti-*Giardia* SIgA and *Giardia* detected in matched stool samples, consistent with prior studies that compared mean salivary SIgA for individuals with microscopy-detected *Giardia* in matched stool samples to those without,[43,44] and weaker associations between anti-*C. jejuni* SIgA and *Campylobacter* in matched stool samples. The results from our threshold analyses indicated limited immunopositive pathogen-specific SIgA responses when those pathogens were detected in matched stool samples. There are biologically plausible reasons for these results. One reason, is that high rates of asymptomatic infections have been documented for both *Giardia* and *C. jejuni*,[1,2,45] and asymptomatic infections may result in a lack of an elevated immune response even though the pathogens were detected in matched stool samples. Another reason, is that SIgA for specific pathogens may be elevated for longer periods than shedding of those pathogens in stool, potentially leading to higher population mean pathogen-specific SIgA concentrations in children whose stool samples were negative for those pathogens and resulting in a higher immunopositive threshold. For example, for *C. jejuni* elevated IgA responses can last for up to 11 weeks after infection,[46] but shedding of *C. jejuni* has been

documented to last as little as two to three or two to ten weeks.[47,48] It is also important to note, that we measured anti-*C. jejuni* SIgA whereas the *Campylobacter* detection in stool included *C. jejuni*, *C. coli*, and *C. lari*. However, because we used serum assays and adapted them for use with saliva, we cannot discern whether these results are due to biological mechanisms or poor assay performance, since the assays are not validated for use with saliva. While we adapted the dilutions used in the assays to account for lower SIgA concentrations in saliva compared to serum,[44,49] there are other potential issues with the use of saliva in serum assays, including the differences in composition between the two matrices leading to possible interference with the ELISA assays when using saliva.[50]

The findings from our study need to be interpreted with its limitations in mind. We used a cross-sectional study design so were not able to make comparisons of salivary SIgA between the same population of children at different time points. We found high variability in our total SIgA estimates and the negative association between total SIgA and the number of concurrent infections was no longer significant after removing outliers. We were limited by sample size for children with no detected enteric infections in matched stool samples, given the high prevalence of any enteric infection (infection with ≥1 enteric pathogen) in the parent study, especially among children older than one year.[30] Approximately two-thirds of our study population was for children aged 1-2 years, with a smaller sample of children aged 3-6 years. The challenge associated with high variability of SIgA concentrations in whole saliva between- and within-individuals has previously been documented.[36] There are a number of external factors that could affect salivary SIgA concentrations that we were not able to control for in our model. These factors include psychological stress,[51] diurnal variations,[52] child behavior like restlessness and crying [53] and dental health.[54] It is also important to note that we defined infections with enteric pathogens as those detected by the GPP in the MapSan analysis of matched stool samples. Children in our cohort may have been experiencing infections with other pathogens not included in the GPP, recent infections where pathogens had not yet begun shedding in stool and thus were not detectable in the matched stool samples,

past infections with continued shedding but no active symptoms, or passage of pathogens detected in the GPPs that did not result in active infection.

In summary, we found that children in a low-income urban slum experiencing high prevalence of enteric infections did not show differences in salivary SIgA with age found in high-income countries, and may be experiencing a suppression of immunological development during early life stages. Particularly notable were the lower total salivary SIgA concentrations with increasing numbers of concurrent infections combined with similar concentrations of total salivary SIgA between infants and pre-school aged children, a life stage where the immune system is developing and where SIgA levels are expected to increase. We also found some evidence of associations between anti-*Giardia and* anti-*C. jejuni* SIgA in saliva and the detection of those pathogens in matched stool samples, but evidence of immunopositive immune responses was limited. Due to the limitations of salivary SIgA diagnostics, including high variability of SIgA levels and a number of external factors we were not able to control for in our analyses, as well as the methodological issues with the pathogen-specific ELISA assays, our findings need to be interpreted with caution. Future directions for this research could include sampling the same children at multiple time points to analyze changes in salivary SIgA over time, in a setting where infections with enteric pathogens are common. Other next steps for this research could include further exploration of pathogen-specific salivary SIgA as biomarkers to estimate specific infections,[55] and compare those to enteric pathogen detection in matched stool samples, a research need that has also been outlined by researchers at the US EPA.[56]

5.6 References

1 Kotloff KL, Nataro JP, Blackwelder WC, *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 2013; **382**: 209–22.

2 Platts-Mills JA, Babji S, Bodhidatta L, *et al.* Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet Glob Heal* 2015; **3**: e564–75.

3 GBD. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis* 2018; **18**: 1211–28.

4 Kosek MN, MAL-ED Network Investigators TM-EN. Causal Pathways from Enteropathogens to Environmental Enteropathy: Findings from the MAL-ED Birth Cohort Study. *EBioMedicine* 2017; **18**: 109–17.

5 Crane RJ, Kelsey JDJ, Berkley JA. Environmental enteric dysfunction: An Overview. *Food Nutr Bull* 2015; **36**: 576–87.

6 Harper KM, Mutasa M, Prendergast AJ, Humphrey J, Manges AR. Environmental enteric dysfunction pathways and child stunting: A systematic review. *PLoS Negl Trop Dis* 2018; **12**: e0006205.

7 UNICEF, WHO WBG. Levels and trends in child malnutrition. 2018 https://www.who.int/nutgrowthdb/2018-jme-brochure.pdf (accessed June 13, 2019).

8 Binnicker MJ. Multiplex Molecular Panels for Diagnosis of Gastrointestinal Infection: Performance, Result Interpretation, and Cost-Effectiveness. *J Clin Microbiol* 2015; **53**: 3723–8.

9 Jimenez L, Duggan CP. Biomarkers of Environmental Enteric Dysfunction: The Good, the Bad, and the Ugly. *J Pediatr Gastroenterol Nutr* 2017; **65**: 4–5.

10 Brown J, Cumming O. Stool-Based Pathogen Detection Offers Advantages as an Outcome Measure for Water, Sanitation, and Hygiene Trials. *Am J Trop Med Hyg* 2019; **0**: 1–2.

11 Arnold BF, Scobie HM, Priest JW, Lammie PJ. Integrated serologic surveillance of population immunity and disease transmission. *Emerg Infect Dis* 2018; **24**: 1188–94.

12 McMurtry CM, Pillai Riddell R, Taddio A, *et al.* Far From "Just a Poke". *Clin J Pain* 2015; **31**: S3–11.

13 Beltrami EM, Williams IT, Shapiro CN, Chamberland ME. Risk and Management of Blood-Borne Infections in Health Care Workers. *Clin Microbiol Rev* 2000; **13**: 385–407.

14 Campisi G, Di Fede O, Roccia P, Di Nicola F, Falaschinp S, Lo Muzio L. Saliva: its value as a biological matrix and current methods of sampling. 2006 https://journals.sagepub.com/doi/pdf/10.1177/1721727X0600400102 (accessed July 2, 2019).

15 Koni AC, Scott RA, Wang G, *et al.* DNA yield and quality of saliva samples and suitability for large-scale epidemiological studies in children. *Int J Obes* 2011; **35**: S113–8.

16 Zabriskie JB. Essential Clinical Immunology. Cambridge, UK: Cambridge University Press, 2009 http://sacema.org/uploads/Essential-Clinical-Immunology.pdf (accessed July 5, 2019).

17 Brandtzaeg P. Do Salivary Antibodies Reliably Reflect Both Mucosal and Systemic Immunity? *Ann N Y Acad Sci* 2007; **1098**: 288–311.

18 Tsujita S, Morimoto K. Secretory IgA in saliva can be a useful stress marker. *Environ Health Prev Med* 1999; **4**: 1–8.

19 Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. Proc. R. Soc. B. 2015; **282**. DOI:10.1098/rspb.2014.3085.

20 Fitzsimmons SP, Evans MK, Pearce CL, Sheridan MJ, Wientzen R, Cole MF. Immunoglobulin A subclasses in infants' saliva and in saliva and milk from their mothers. *J Pediatr* 1994; **124**: 566–73.

21 Ben-Aryeh H, Fisher M, Szargel R, Laufer D. Composition of whole unstimulated saliva of healthy children: Changes with age. *Arch Oral Biol* 1990; **35**: 929–31.

22 Mantis NJ, Forbes SJ. Secretory IgA: arresting microbial pathogens at epithelial borders. *Immunol Invest* 2010; **39**: 383–406.

23 Wells JM, Brummer RJ, Derrien M, *et al.* Homeostasis of the gut barrier and potential biomarkers. *Am J Physiol Liver Physiol* 2017; **312**: G171–93.

24 Mantis NJ, Rol N, Corth B. Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. *Mucosal Immunol* 2011; **4**: 603–11.

25 Rojas R, Apodaca G. Immunoglobulin transport across polarized epithelial cells. *Nat Rev Mol Cell Biol* 2002; **3**: 944–56.

26 Kaetzel CS, Robinson JK, Chintalacharuvu KR, Vaerman J-P, Lamm ME. The polymeric immunoglobulin receptor (secretory component) mediates transport of immune complexes across epithelial cells: A local defense function for IgA (epitheal mncous/nual immuniy). *Proc Nati Acad Sci* 1991; **88**: 8796–800.

27 Brown J, Cumming O, Bartram J, *et al.* A controlled, before-and-after trial of an urban sanitation intervention to reduce enteric infections in children: research protocol for the Maputo Sanitation (MapSan) study, Mozambique. *BMJ Open* 2015; **5**. DOI:10.1136/bmjopen-2015-008215.

28 Knee J, Sumner T, Adriano Z, *et al.* Risk factors for childhood enteric infection in urban Maputo, Mozambique: A cross-sectional study. *PLoS Negl Trop Dis* 2018; **12**: e0006956.

29 Egorov AI, Griffin SM, Ward HD, Reilly K, Fout GS, Wade TJ. Application of a salivary immunoassay in a prospective community study of waterborne infections. *Water Res* 2018; **142**: 289–300.

30 Knee J. Quantifying the impact of an urban onsite shared sanitation intervention on child health in Maputo, Mozambique: the Mapsan trial. 2019.

31 Lehmann FS, Burri E, Beglinger C. The role and utility of faecal markers in inflammatory bowel disease. *Therap Adv Gastroenterol* 2015; **8**: 23–36.

32 de Serres F, Blanco I. Role of alpha-1 antitrypsin in human health and disease. *J Intern Med* 2014; **276**: 311–35.

33 Murr C, Widner B, Wirleitner B, Fuchs D. Neopterin as a Marker for Immune System Activation. *Curr Drug Metab* 2005; **3**: 175–87.

34 Sonesson M, Hamberg K, Lundin Wallengren M-L, Matsson L, Ericson D. Salivary IgA in minor-gland saliva of children, adolescents, and young adults. *Eur J Oral Sci* 2011; **119**: 15–20.

35 Eliasson L, Birkhed D, Österberg T, Carlén A. Minor salivary gland secretion rates and immunoglobulin A in adults and the elderly. *Eur J Oral Sci* 2006; **114**: 494–9.

36 Brandtzaeg P. Secretory immunity with special reference to the oral cavity. *J Oral Microbiol*

2013; **5**. DOI:10.3402/jom.v5i0.20401.

37 Platts-Mills J, Liu J, Houpt E. New concepts in diagnostics for infectious diarrhea. *Mucosal Immunol* 2013; **6**. DOI:10.1038/mi.2013.50.

38 Katona P, Katona-Apte J. The Interaction between Nutrition and Infection. *Clin Infect Dis* 2008; **46**: 1582–8.

39 Platts-Mills JA, Taniuchi M, Uddin MJ, *et al.* Association between enteropathogens and malnutrition in children aged 6-23 mo in Bangladesh: A case-control study. *Am J Clin Nutr* 2017; **105**: 1132–8.

40 Watson RR, McMurray DN, Martin P, Reyes MA. Effect of age, malnutrition and renutrition on free secretory component and IgA in secretions. *Am J Clin Nutr* 1985; **42**: 281–8.

41 Fagerås M, Tomičić S, Voor T, Björkstén B, Jenmalm MC. Slow Salivary Secretory IgA Maturation May Relate to Low Microbial Pressure and Allergic Symptoms in Sensitized Children. *Pediatr Res* 2011; **70**: 572–7.

42 Weemaes C, Klasen I, Göertz J, Beldhuis-Valkis M, Olafssonz O, Haraldsson A. Development of Immunoglobulin A in Infancy and Childhood. *Scand J Immunol* 2003; **58**: 642–8.

43 Rodríguez OL, Isabel Hagel, González Y, *et al.* Secretory IgA Antibody Responses in Venezuelan Children Infected with Giardia duodenalis. *J Trop Pediatr* 2004; **50**: 68–72.

44 Saad El-Gebaly NM, Fawzy Halawa E, Ezzat Moussa HM, Rabia I, Abu-Zekry M. Saliva and sera IgA and IgG in Egyptian Giardia-infected children. *Parasitol Res* 2012; **111**: 571–5.

45 Platts-Mills JA, Kosek M. Update on the burden of Campylobacter in developing countries. *Curr Opin Infect Dis* 2014; **27**: 444–50.

46 Aagaard Strid M, Engberg J, Brandt Larsen L, Begtrup K, Mølbak K, Angeliki Krogfelt K. Antibody Responses to Campylobacter Infections Determined by an Enzyme-Linked Immunosorbent Assay: 2-Year Follow-Up Study of 210 Patients. *Clin Diagn Lab Immunol* 2001; **8**: 314–9.

47 World Health Organization. Foodborne disease outbreaks: Guidelines for investigation and control. Geneva, 2008 http://apps.who.int/iris/bitstream/handle/10665/43771/9789241547222_eng.pdf;jsessionid=23A76E49C434117ABE9923C77139681F?sequence=1 (accessed Aug 21, 2018).

48 Kapperud G, Lassen J, Ostroff SM, *et al.* Clinical Features of Sporadic Campylobacter Infections in Norway. *Scand J Infect Dis* 1992; **24**: 741–9.

49 Friedman MG, Entin N, Zedaka R, Dagan R. Subclasses of IgA antibodies in serum and saliva samples of newborns and infants immunized against rotavirus. *Clin Exp Immunol* 1996; **103**: 206–11.

50 Jaedicke KM, Taylor JJ, Preshaw PM. Validation and quality control of ELISAs for the use with human saliva samples. *J Immunol Methods* 2012; **377**: 62–5.

51 Engeland CG, Hugo FN, Hilgert JB, *et al.* The effects of psychological distress on salivary secretory immunity HHS Public Access. *Brain Behav Immun* 2016; **52**: 11–7.

52 Lim PW, Nambiar S, Muhardi L, Abdul Kader UH, Garssen J, Sandalova E. Young Children Display Diurnal Patterns of Salivary IgA and Alpha-Amylase Expression Which Are Independent of Food Intake and Demographic Factors. *Biomed Res Int* 2019; **2019**: 3687416.

53 Tenovuo J, Lehtonen O-PJ, Aaltonen AS, Vilja P, Tuohimaa4 P. Antimicrobial Factors in Whole Saliva of Human Infants. 1986 http://iai.asm.org/ (accessed June 28, 2019).

54 Marcotte H, Lavoie MC. Oral Microbial Ecology and the Role of Salivary Immunoglobulin A. 1998 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC98907/pdf/mr000071.pdf (accessed July 5, 2019).

55 Exum NG, Pisanic N, Granger; DA, *et al.* Use of Pathogen-Specific Antibody Biomarkers to Estimate Waterborne Infections in Population-Based Settings. *Curr Environ Heal Reports* 2016; **3**: 322–34.

56 Wade TJ, Augustine SAJ, Griffin SM, *et al.* Asymptomatic norovirus infection associated with swimming at a tropical beach: A prospective cohort study. *PLoS One* 2018; **13**: 1–14.

6 **Chapter 6:** Summary, implications, future research and conclusion

6.1 Summary of findings

To quantify associations between proxy measures of external enteric exposure and adverse child health outcomes, we compiled over 90,000 individual participant data points from twenty studies across a range of low- and middle-income settings for the first research aim. We found strong evidence of an association between household-level fecal contamination in drinking water and both diarrhea and impaired linear growth in children. Our analyses also implicated contaminated child hands in diarrhea and contaminated fomites in linear growth. The study highlighted the paucity of evidence along common exposure pathways other than drinking water, especially food and soil. Secondary analyses found differences in effect sizes by child age and between rural and urban settings, although data for urban areas was limited to drinking water and diarrhea. Secondary analyses also indicated that diarrhea and linear growth burdens were driven by high levels of fecal contamination. While this finding could be interpreted by suggesting low levels of fecal contamination have lower overall adverse effects on child health outcomes, it could also have resulted from misclassification of exposure due to limits of detections at low FIB concentrations.

For the second research aim we set out to evaluate these proxy measures of external exposure by adapting an exposure measurement error framework previously published for time-series studies of air pollution epidemiology, to estimate whether enteric exposure proxies used in the analyses of the first aim bias true fecal exposure-health outcome associations. We employed the extensive drinking water quality and child health datasets compiled for the first aim and found that household-level exposure assessments that do not consider potential exposure in the community may attenuate true water quality-diarrhea associations, particularly for older children. We also found that using single water quality samples that do not account for temporal variability in water quality, instead of an aggregate of multiple longitudinally collected samples, may attenuate the true effects of water quality on child growth.

We considered an emerging measure of internal exposure for the third aim, by comparing salivary antibody levels in saliva to enteric pathogens and EED biomarkers found in matched stool samples in a sample of children enrolled in the MapSan trial in Maputo, Mozambique. We found that total non-specific salivary SIgA was lower in children experiencing a higher number of enteric infections, and found no association between salivary SIgA and child age. Taken together, these findings may indicate that high prevalence of enteric infection can lead to a suppression of immunological development in children aged one to six years, a life stage where the immune system is developing and where SIgA levels are expected to increase. We also found limited association between pathogen-specific salivary SIgA for *C. jejuni and Giardia* with the detection of *Campylobacter and Giar*dia in matched stool samples.

6.1.1 Limitations

While much of this dissertation was intended to describe the deficiencies of current approaches for measuring exposure to enteric pathogens and explore potential alternatives, the first research aim employed data generated precisely from these approaches and as a systematic review relied on previous research. Although we identified an association between certain exposure measures and child health, the data was largely limited to the drinking water pathway. Moreover, our assessment of the potential for measurement error for the second research aim suggests that these results may underestimate the true relationship. As such, except for select secondary subgroup analyses, we were limited to drawing general conclusions on the associations between fecal contamination and health. We were not able to make more specific contributions, such as whether in certain areas or regions select pathways of exposure are more dominant compared to others, or whether pathway-specific exposure-outcome associations are modified by overall levels of exposure experienced by individuals. None of the included studies measured fecal contamination along all six pathways included in the analyses, so we were not able to model all pathways simultaneously to tease apart the relative contributions from different pathways to child diarrhea and impaired linear growth. In addition, our findings need to be interpreted with caution due to risk of uncontrolled confounding in our analyses. The limited data along some common transmission pathways,

such as soil and food, meant we were limited in the conclusions we could draw on the associations between fecal contamination along all six included pathways and diarrhea and growth.

The second aim was limited to evaluating two potential sources of random exposure measurement error in external enteric exposure assessments. There are many other potential sources of error, such as assigning household- or community-level exposure measures to individuals and environmental sample processing errors (i.e. during sample collection, transport, or analyses) that can lead to under- or over-estimates of concentrations of enteric pathogens or fecal indicator organisms in different environmental reservoirs or to false negative or false positive results. We also did not extend this study beyond the drinking water pathway, because we were limited by data availability for the other common fecal-oral transmissions pathways. Each pathway has its own sampling decisions that need to be made when designing an exposure assessment, such as the location, frequency of sampling and sample type, which can all influence errors in estimating exposure differentially.

The third aim was limited to a cross-sectional sample of children for whom we processed saliva samples. Accordingly, we were limited to considering differences between groups in our analyses and were not able to investigate changes in salivary antibody levels over time. Analyzing samples of the same children longitudinally over time would have been especially valuable due to our unexpected findings of consistent salivary SIgA concentrations between children of different age groups. Given the infancy of salivary antibody diagnostics, we were also limited by the conclusions we could draw from our pathogen-specific SIgA assays. More validation of the ELISA assays designed for use with serum or plasma is needed before the limited association between anti-*C. jejuni* and anti-*Giardia* salivary antibodies and positive PCR results for these pathogens in matched stool samples can be confirmed.

6.2 Implications

Despite uncertainty from individual studies and prior reviews, the findings from the first research aim confirmed the fundamental association between fecal contamination at the household level and diarrhea. The study also provided evidence of the effects of fecal contamination on impaired linear growth, a

hypothesized outcome that recent trials designed to improve health by reducing fecal contamination have been unable to confirm. These results suggest that the inability of WaSH interventions to achieve consistent improvements in measured health outcomes may be due to their failure to adequately reduce fecal contamination. However, policy makers and program implementers may question the value of WaSH infrastructure investments to reduce fecal contamination and subsequently improve child health given the relatively small effect sizes we found in this study. For example, for children experiencing 1-$\log_{10}$ higher FIB concentrations in drinking water, i.e. an order of magnitude more fecal contamination, the odds of diarrhea were 1.09 (95% CI: 1.04, 1.13) and the difference in HAZ score was -0.04 (95% CI: -0.06, -0.01). Therefore, there may be questions about the biological importance of fecal contamination as it pertains to the measured health outcomes when using these proxy measures of external exposure to estimate enteric exposure-health outcome relationships.

To address these questions and provide more context to the findings from the first aim, the second aim demonstrated how measurement error frameworks can be used to better understand the true contribution of fecal contamination to adverse child health outcomes. Our analyses found substantial regression dilution bias from just two potential sources of error, suggesting that the proxy measures of external exposure used in the first aim may have suffered from random exposure measurement error and the true exposure-outcome relationships may be stronger than what the findings from our analyses suggested. The broader contribution of this study is an initial effort to apply an exposure measurement error framework to questions surrounding methods of external enteric exposure. Approaching the design of exposure assessments with a perspective of limiting potential sources of measurement error can help design enteric pathogen exposure assessments that are more biologically relevant to health outcomes of interest.

The third aim of this dissertation considered antibody concentrations in saliva as a potential measure of past exposure to enteric pathogens. Implications from this study include that saliva may be a promising biological matrix based on the relative ease of sample collection in a resource-constrained field setting, especially compared to collecting stool samples. However, questions remain about the utility of salivary

antibodies to characterize gastrointestinal health and past exposure to enteric pathogens. The challenges associated with antibody detection in saliva mean more assay development and validation are still needed before salivary antibodies can be measured reliably at scale to estimate past exposure to enteric pathogens.

6.3 Future research

There is a need to dedicate resources and research to characterizing environmental exposures in underserved and underrepresented populations. Advanced methods to estimate enteric pathogen exposure are needed to inform the design of improved interventions to reduce infectious disease burdens, particularly among young children living in low-income countries. There is an opportunity to move enteric exposure science forward through cross-sectoral collaboration that combines modeling, observational, microbiological, epidemiological and social/behavioral tools for improved measures of exposure. Such improved measures might enable the WaSH sector to "fail-fast" by evaluating interventions based on whether they reduce exposure to enteric pathogens instead of more distal and perhaps more slow-developing health outcomes. Improved measures of exposure could also be used for results-based financing of intervention implementation, i.e. if implementers were paid for performance, and defining performance as reducing exposure, instead of being rewarded for increasing coverage for WaSH services with no regard to whether or not they actually reduce exposure.

Best practices for microbiological and observational methods used for exposure assessments need to be standardized to enable better comparisons across different studies and settings. Developing a database framework with uniform data reporting standards to allow for pooling of data across multiple studies would allow for better evaluation and generalizability of strengths and weaknesses of exposure methods. The IPD analyses conducted for this dissertation demonstrated the value of pooling data from a number of studies to quantify exposure-outcome relationships, but further research could explore opportunities to continue to pool data to estimate other common exposure-response relationships, including neglected fecal-oral transmission pathways such as food and soil, and aerosolized pathogens (an additional pathway

now getting more attention), to provide a more complete picture of exposure along different transmission pathways.

The IPD analyses were limited to using indicators of fecal contamination as proxies for enteric pathogen exposure, largely because environmental pathogen data are limited. However, emerging pathogen detection methods are making it possible to detect multiple enteric pathogens in environmental samples at scale. The Taqman Array Card (TAC), a real-time multiplex PCR assay, can detect multiple enteric pathogens simultaneously encompassing viral, bacterial, protozoal, and helminth targets for enteric infections.[1] Examples of recent environmental applications of the TAC method include the simultaneous detection of a number of enteric pathogens in surface water, soil and infant weaning food in Kenya[2,3]. One limitation of PCR-based methods is the need to pre-specify targets, which can be especially challenging for pathogen detection given the potential diversity of pathogens in the environment. Environmental metagenomics, the sequencing and analysis of all DNA in environmental samples, can circumvent this problem.[4] For example, metagenomic approaches were recently used to distinguish between foodborne disease outbreak strains of *Salmonella*[5], and to identify the likely causes of diarrheagenic *E. coli* in Ecuador[6]. Limitations of metagenomics include poor sensitivity if enteric pathogens are at low prevalence in the microbial community, high cost, required bioinformatics expertise, and the need for improved analysis pipelines for identifying pathogens.

Instead of using laboratory intensive approaches to estimate pathogen occurrence in the environment, there is room for innovation to detect enteric pathogens on-site in environmental compartments by using biosensing technologies that are in development.[7] Such methods could be used to rapidly detect pathogens in different environmental reservoirs and thus provide more complete data on enteric pathogen concentrations along different fecal-oral transmission pathways. If developed further, biosensing methods have the advantage that they can be deployed in resource-constrained and remote settings, because they remove the need for sample collection, transport, and intensive laboratory processing. As a result, the

potential of these tools reach beyond generating research data, and could empower local stakeholders to conduct independent environmental surveillance.

Questions that this dissertation was not able to comprehensively address due to the lack of pathogen-specific environmental data and because few studies have measured exposure via multiple transmission pathways simultaneously, pertain to the teasing apart of the (likely context-specific) dominant risks of exposure. Specific future research questions include:

- What are the primary modes of enteric pathogen transmission?
- What are the relative contributions to enteric health burdens from different modes of transmission?

As more environmental data on pathogen occurrence becomes available, the interpretation and statistical analysis of these data will need to become a research priority to quantify enteric pathogen-infectious disease outcome relationships. In addition, research on the temporal and spatial variability of enteric pathogens in environmental matrices is needed to inform sampling strategies. The exposure measurement error simulations conducted for this dissertation, demonstrated the potential value of approaching exposure assessments from a perspective of biological relevance as it pertains to infectious disease health outcomes. There is a need for longitudinal studies to provide rigorous evidence on relative contributions from specific transmission pathways and pathogens to adverse health outcomes, and the design of those studies should employ exposure measurement error frameworks to reduce error between measured and true exposure and design more biologically-relevant exposure assessments.

Since the second aim of this dissertation was limited to considering just two possible components of error and did not evaluate how to practical it is to reduce these errors, future research questions include:

- Do other components of error introduce exposure measurement error in to health effect estimates?
- How cost-effective is attempting to reduce these errors in enteric exposure assessments?

To inform external exposure assessment design, measures of internal exposure can be useful since they provide data on past exposure to help frame external exposure assessments. Characterizing internal exposure using stool samples can be resource intensive, especially for epidemiological studies in resource-constrained settings,[8] so there is opportunity for further research with alternative biological matrices. We piloted the use of salivary antibodies with singleplex immunoassays for this dissertation, and compared salivary SIgA concentrations to enteric infections found in stool. More method development and validation is needed to use saliva in immunoassays at scale. The US EPA is at the forefront of recent methodological developments to test for salivary antibody response to multiple specific pathogens simultaneously using Luminex-based immunoassays.[9,10] Another approach seeing increased attention as a way to avoid collecting individual stool samples, is fecal sludge surveillance. Most prominently, sewage is used for poliovirus detection in global eradication efforts,[11] but it has also seen application in advanced warning of viral outbreaks.[12,13] Fecal sludge is attractive because it is non-invasive, logistically easier to collect than stool samples and can serve as a composite sample of an entire community's feces and can thus indicate past exposure at the community level.

Since the third aim of this dissertation was limited to measuring salivary IgA in a cross-sectional sample of children, there are future research questions that came out of this aim that follow-up studies might address by collecting longitudinal samples and measuring different salivary biomarkers:

- How does salivary IgA change over time in children experiencing high prevalence of enteric infection?
- How does salivary IgA compare to other biomarkers of antibody response?

6.4 Conclusion

This dissertation sought to answer questions surrounding the characterization of exposure to enteric pathogens, with a focus on populations that are most commonly exposed and most vulnerable to these exposures: young children in low-income countries. We approached these questions from various angles,

drawing on different disciplines of environmental health. We started by compiling a large dataset of secondary data to estimate enteric exposure-child health outcome relationships, finding that commonly used proxies of exposure are associated with child diarrhea and linear growth although the strength of the associations as well as the evidence available varied by fecal-oral transmission pathway. We critically evaluated those findings by employing statistical measurement error frameworks, finding that these proxy measures of exposure may be suffering from substantial regression dilution bias, attenuating true exposure-outcome relationships towards no effect. This dissertation also included bench research at the Georgia Institute of Technology to pilot the use of salivary antibodies as an alternative biological matrix for internal exposure assessments, indicating that concurrent enteric infections may be suppressing immune system development in young children, although further validation of salivary diagnostics is needed to confirm this result. The findings from each of the research aims contribute to the scientific endeavor of estimating enteric exposure and its effect on adverse child health outcomes. Each aim also underlines opportunities for further progress in this area of research. These opportunities along with others highlighted by an interdisciplinary working group of biomarker researchers, microbiologists, exposure scientists, environmental engineers, and epidemiologists are being published as part of this dissertation, to provide a review and recommendations on enteric exposure methods to move this field of research forward.

6.5 References

1 Liu J, Gratz J, Amour C, *et al.* A laboratory-developed taqman array card for simultaneous detection of 19 enteropathogens. *J Clin Microbiol* 2013; **51**: 472–80.

2 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**: 10263–74.

3 Tsai K, Simiyu S, Mumma J, *et al.* Enteric Pathogen Diversity in Infant Foods in Low-Income Neighborhoods of Kisumu, Kenya. *Int J Environ Res Public Health* 2019; **16**: 506.

4 Tringe SG, Rubin EM. Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* 2005; **6**: 805–14.

5 Huang AD, Luo C, Pena-Gonzalez A, Weigand MR, Tarr CL, Konstantinidis KT. Metagenomics of two severe foodborne outbreaks provides diagnostic signatures and signs of coinfection not attainable by traditional methods. *Appl Environ Microbiol* 2017; **83**: e02577-16.

6 Peña-Gonzalez A, Soto-Girón MJ, Smith S, *et al.* Metagenomic Signatures of Gut Infections Caused by Different Escherichia coli Pathotypes. *Appl Environ Microbiol* 2019; **85**: e01820-19.

7 Korostynska O, Mason A, Al-Shamma'a AI. Monitoring pollutants in wastewater: Traditional lab based versus modern real-time approaches. In: Smart Sensors, Measurement and Instrumentation. Springer International Publishing, 2013: 1–24.

8 Brown J, Cumming O. Stool-Based Pathogen Detection Offers Advantages as an Outcome Measure for Water, Sanitation, and Hygiene Trials. *Am J Trop Med Hyg* 2019; **0**: 1–2.

9 Griffin SM, Chen IM, Fout GS, Wade TJ, Egorov AI. Development of a multiplex microsphere immunoassay for the quantitation of salivary antibody responses to selected waterborne pathogens. *J Immunol Methods* 2011; **364**: 83–93.

10 Augustine SAJ, Eason TN, Simmons KJ, *et al.* Developing a Salivary Antibody Multiplex Immunoassay to Measure Human Exposure to Environmental Pathogens. *J Vis Exp* 2016. DOI:10.3791/54415.

11 Hovi T, Schulman LM, Van der Avoort H, Deshpande J, Roivainen M, De Gourville EM. Role of environmental poliovirus surveillance in global polio eradication and beyond. *Epidemiol Infect* 202AD; **140**: 1–13.

12 Hellmér M, Paxéus N, Magnius L, *et al.* Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Appl Environ Microbiol* 2014; **80**: 6771–81.

13 Smith DB, Paddy JO, Simmonds P. The use of human sewage screening for community surveillance of hepatitis E virus in the UK. *J Med Virol* 2016; **88**: 915–8.

7 Appendices

7.1 **Appendix A –** Research Aim 1: Search Strings

Generic Search String: ((intervention OR programme OR program OR evaluation) AND (wash OR water OR sanitation OR hygiene)) AND ((diarrhea OR diarrhoea OR “diarrheal disease” OR “diarrhoeal disease” OR growth OR anthropometry OR anthropometrics OR HAZ OR LAZ OR “height-for-age” OR “height for age” OR “length-for-age” OR “length for age”) AND (child OR children OR infant))

PubMed: ((intervention[tw] OR programme[tw] OR program[tw] evaluation[tw]) AND (wash[tw] OR water[tw] OR sanitation[tw] OR hygiene[tw])) AND ((diarrhea[tw] OR diarrhoea[tw] OR “diarrheal disease”[tw] OR “diarrhoeal disease”[tw] OR growth[tw] OR anthropometry[tw] OR anthropometrics[tw] OR HAZ[tw] OR LAZ[tw] OR “height-for-age”[tw] OR “height for age”[tw] OR “length-for-age”[tw] OR “length for age”[tw]) AND (child[tw] OR children[tw] OR infant[tw]))

Web of Science: TS = ((intervention OR programme OR program OR evaluation) AND (wash OR water OR sanitation OR hygiene)) AND TS = ((diarrhea OR diarrhoea OR "diarrheal disease" OR "diarrhoeal disease" OR growth OR anthropometry OR anthropometrics OR HAZ OR LAZ OR "height-for-age" OR "height for age" OR "length-for-age" OR "length for age") AND (child OR children OR infant))

EMBASE: ((intervention:ti,ab,kw OR programme:ti,ab,kw OR program:ti,ab,kw OR evaluation:ti,ab,kw) AND (wash:ti,ab,kw OR water:ti,ab,kw OR sanitation:ti,ab,kw OR hygiene:ti,ab,kw)) AND ((diarrhea:ti,ab,kw OR diarrhoea:ti,ab,kw OR ‘diarrheal disease’:ti,ab,kw OR ‘diarrhoeal disease’:ti,ab,kw OR growth:ti,ab,kw OR anthropometry:ti,ab,kw OR anthropometrics:ti,ab,kw OR HAZ:ti,ab,kw OR LAZ:ti,ab,kw OR ‘height-for-age’:ti,ab,kw OR ‘height for age’:ti,ab,kw OR ‘length-for-age’:ti,ab,kw OR ‘length for age’:ti,ab,kw) AND (child:ti,ab,kw OR children:ti,ab,kw OR infant:ti,ab,kw))

7.2 **Appendix B –** Research Aim 1: Supplementary Figures

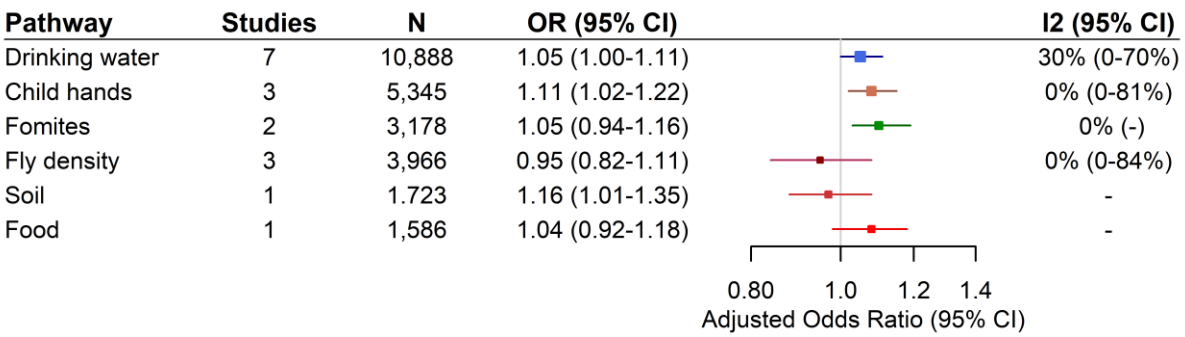


Figure B-1: Odds of stunting for a 1-log$_{10}$ higher median fecal contamination by fecal-oral transmission pathway

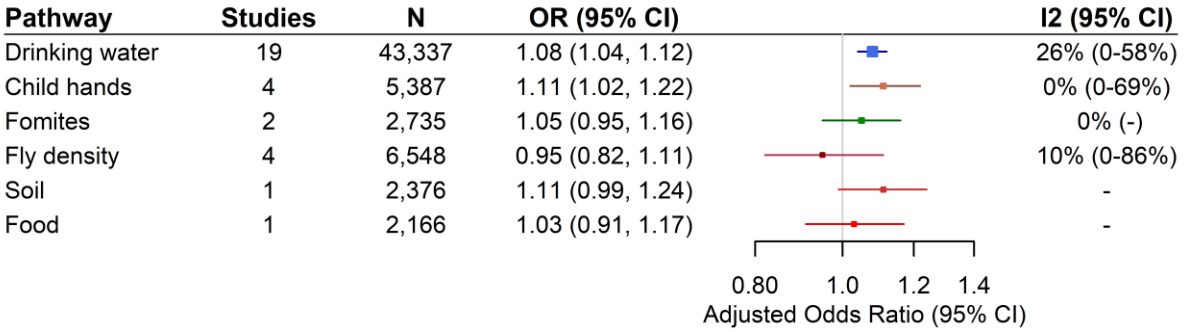


Figure B-2: Odds of diarrhea a 1-log higher fecal indicator bacteria concentrations in drinking water, on child hands, on fomites, in soil, and food, and a 1-log higher in kitchen fly density, using six log$_{10}$ categories (<1, 1-10, 11-100, 101-1,000, 1,001-10,000, 10,000+) instead of four as the explanatory variable.

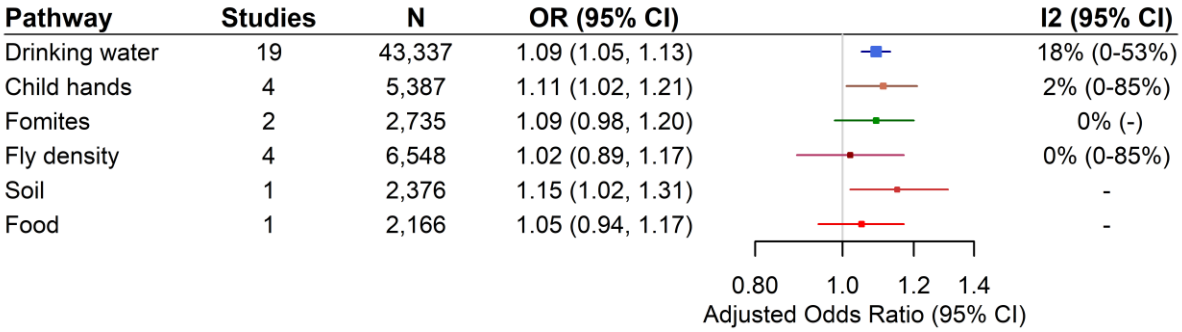


Figure B-3: Odds of diarrhea a 1-log higher fecal indicator bacteria concentrations in drinking water, on child hands, on fomites, in soil, and food, and a 1-log higher kitchen fly density, using a continuous log_{10} transformation as the explanatory variable instead of log_{10} categories.

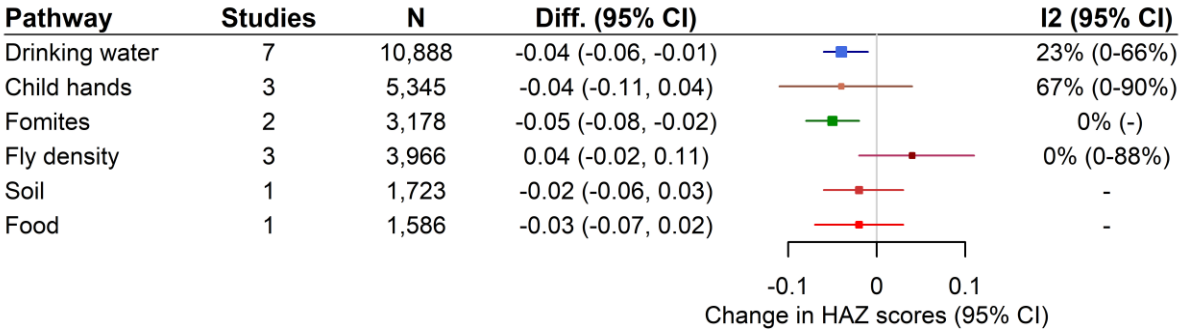


Figure B-4: Difference in height-for-age Z score for 1-log higher median fecal indicator bacteria concentrations in drinking water, on child hands, on fomites, in soil, and food, and 1-log higher median kitchen fly density, using six log_{10} categories (<1, 1-10, 11-100, 101-1,000, 1,001-10,000, 10,000+) instead of four as the explanatory variable.

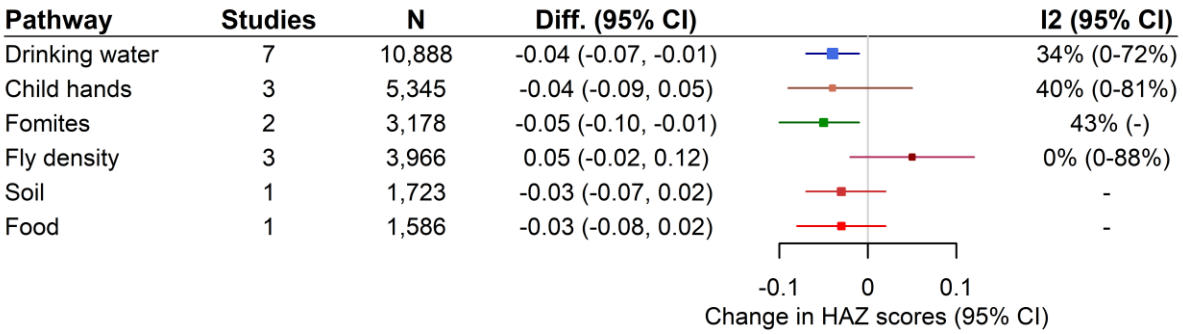
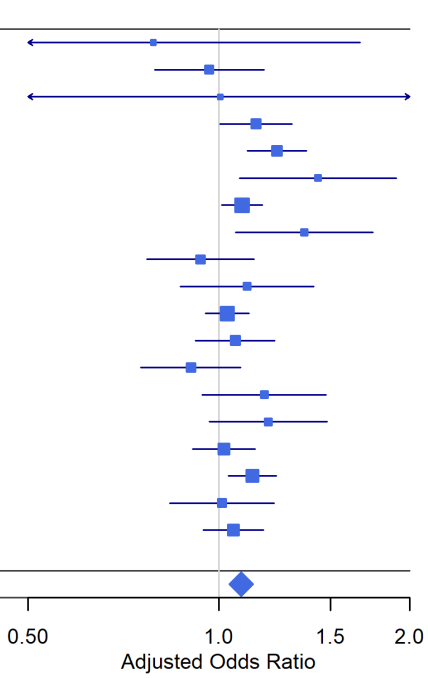


Figure B-5: Difference in height-for-age Z score for 1-log higher median fecal indicator bacteria concentrations in drinking water, on child hands, on fomites, in soil, and food, and 1-log higher median kitchen fly density, using a continuous log_{10} transformation as the explanatory variable instead of log_{10} categories

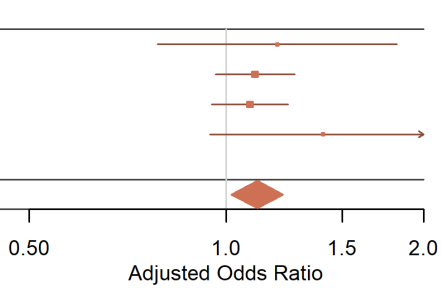
Water

Study	N	OR	95%CI
Arnold, 2010	196	0.79	(0.37 - 1.67)
Benjamin-Chung, 2018	1291	0.97	(0.79 - 1.18)
Boisson, 2010	149	1	(0.43 - 2.35)
Boisson, 2013	5637	1.14	(1 - 1.3)
Brown, 2008	2478	1.23	(1.11 - 1.37)
Clasen, 2005	520	1.43	(1.08 - 1.9)
Clasen, 2014	4948	1.09	(1.01 - 1.17)
Davis, in prep.	4383	1.36	(1.06 - 1.75)
Ercumen, 2015	1895	0.94	(0.77 - 1.14)
Kirby, 2017	580	1.11	(0.87 - 1.41)
Kirby, Nagel, 2019	5519	1.03	(0.95 - 1.12)
Luby, 2015	2530	1.06	(0.92 - 1.22)
Patil, 2015	1293	0.9	(0.75 - 1.08)
Peletz, 2011	347	1.18	(0.94 - 1.48)
Peletz, 2012	1576	1.2	(0.97 - 1.48)
Pickering, Ercumen, 2018	2387	1.02	(0.91 - 1.14)
Pickering, 2019	2563	1.13	(1.04 - 1.23)
Reese, 2019	2297	1.01	(0.84 - 1.22)
Sinharoy, 2017	2748	1.05	(0.95 - 1.17)
Summary	**43337**	**1.09**	**(1.04 - 1.13)**



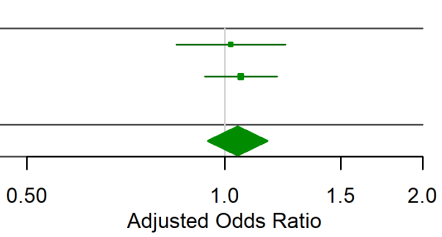
Child hands

Study	N	OR	95%CI
Devamani, 2014	439	1.2	(0.78 - 1.82)
Pickering, Ercumen, 2018	1003	1.11	(0.96 - 1.27)
Pickering, 2019	1752	1.09	(0.95 - 1.24)
Reese, 2019.	517	1.4	(0.94 - 2.09)
Summary	**3711**	**1.11**	**(1.02 - 1.22)**



Fomites

Study	N	OR	95%CI
Benjamin-Chung, 2018 .	1262	1.02	(0.84 - 1.24)
Pickering, 2019	1473	1.06	(0.93 - 1.2)
Summary	**2735**	**1.05**	**(0.94 - 1.16)**



Kitchen fly density

Study	N	OR	95%CI
Benjamin-Chung, 2018	1253	0.78	(0.58 - 1.06)
Clasen, 2014	213	0.82	(0.46 - 1.45)
Pickering, Ercumen, 2018	2355	0.88	(0.48 - 1.58)
Pickering, 2019	2727	1.04	(0.9 - 1.2)
Summary	**6548**	**0.95**	**(0.82 - 1.11)**

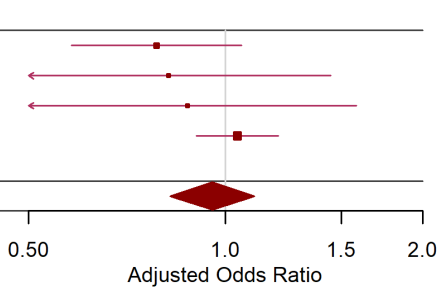
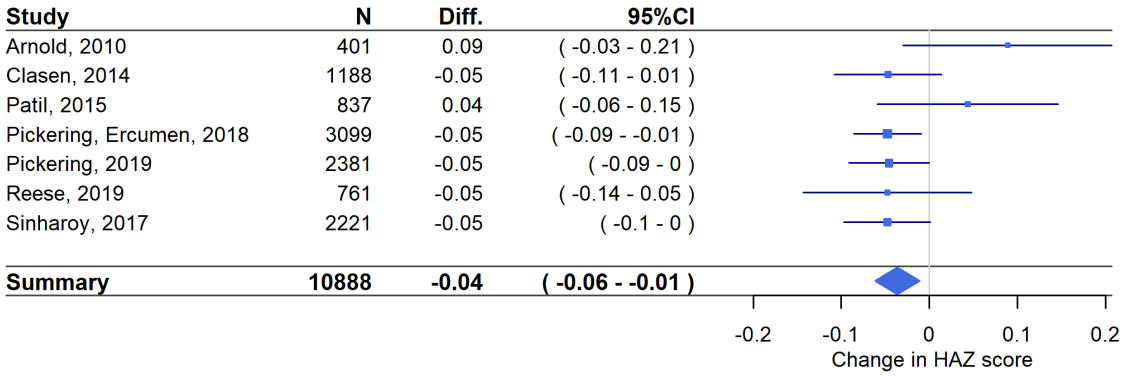
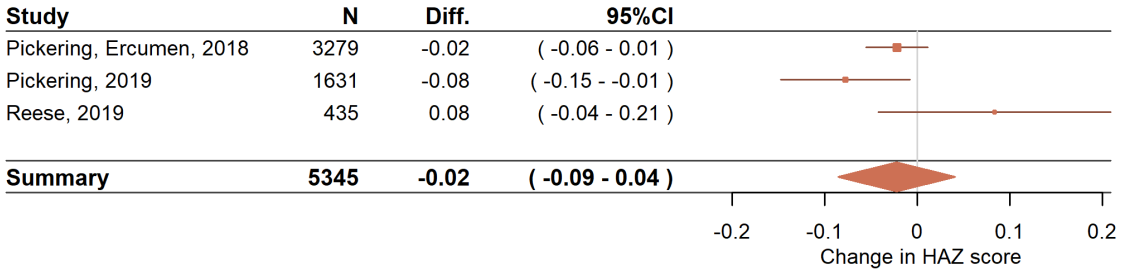


Figure B-6: Odds of diarrhea for 1-$\log_{10}$ higher fecal indicator bacteria concentrations in drinking water, on child hands, and on fomites, and 1-$\log_{10}$ higher kitchen fly density.

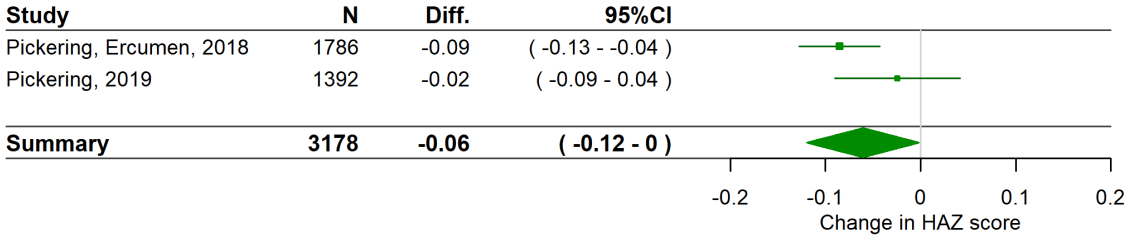
Water



Child hand



Fomites



Kitchen fly density

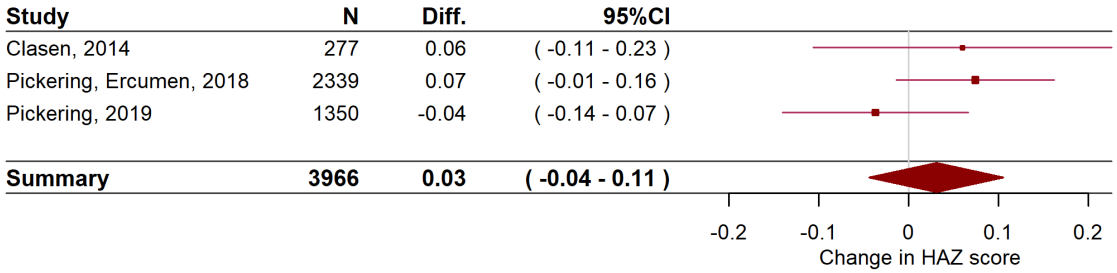
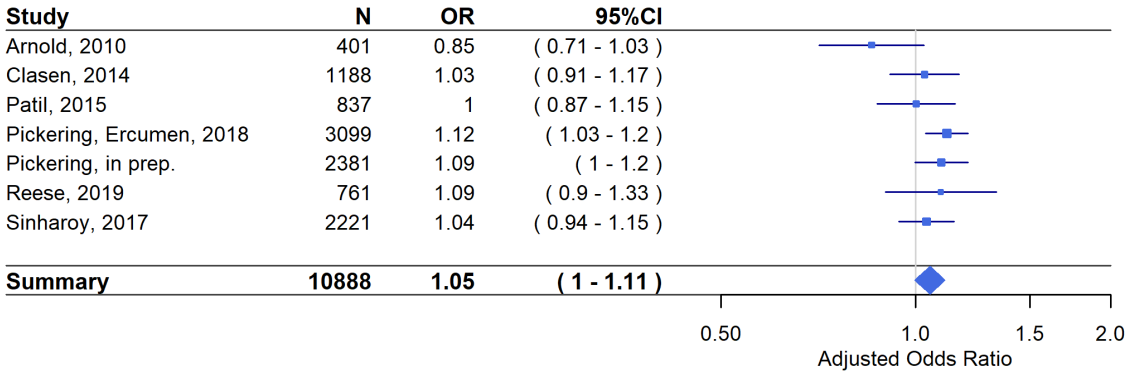
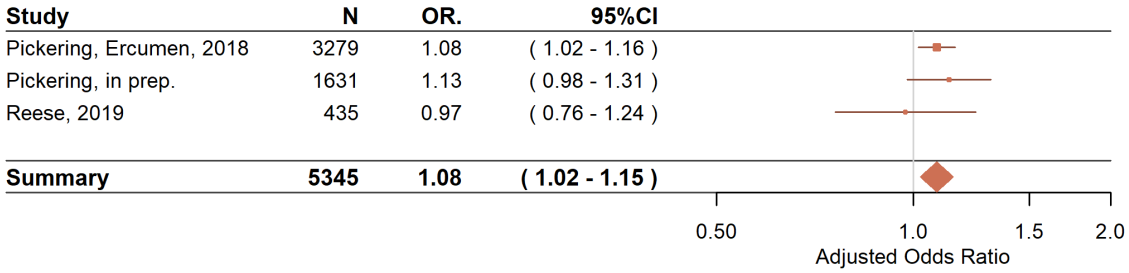


Figure B-7: Difference in HAZ scores for a 1-log_{10} higher median fecal indicator bacteria concentrations in drinking water, on child hands, and on fomites, and a 1-log_{10} higher median kitchen fly density.

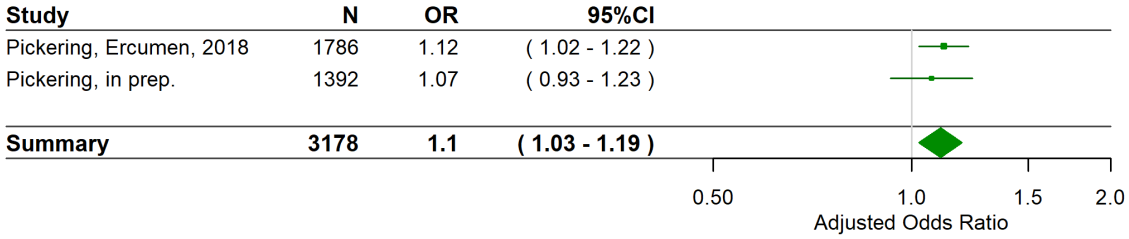
Water



Child hands



Fomites



Kitchen fly density

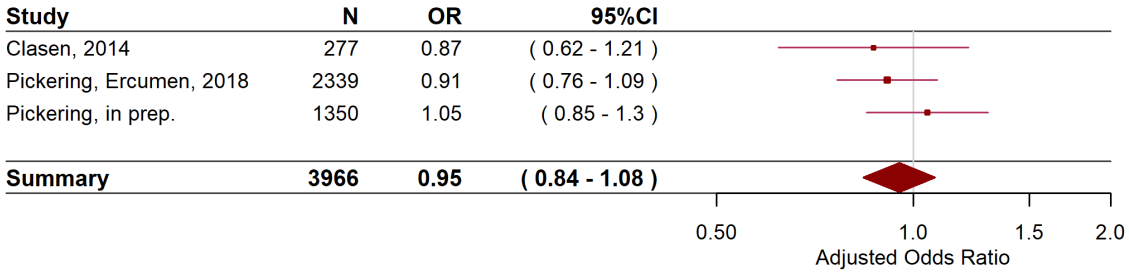


Figure B-8: Odds of stunting for 1-log_{10} higher median fecal indicator bacteria concentrations in drinking water, on child hands, and on fomites, and a 1-log_{10} higher median kitchen fly density.

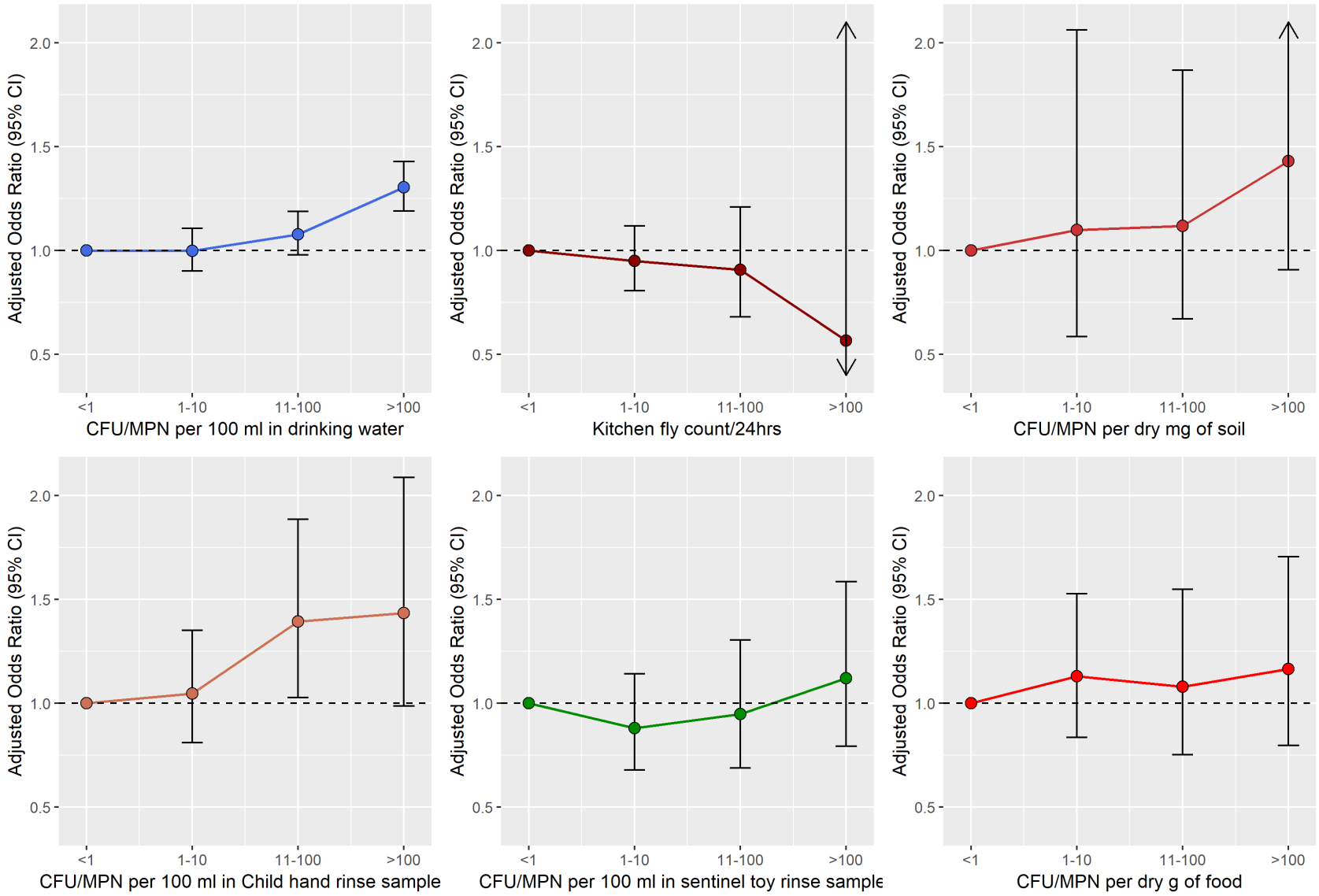


Figure B-9: Odds of diarrhea for 1-log_{10} higher fecal contamination, stratified by log category compared to the reference level of no measured contamination

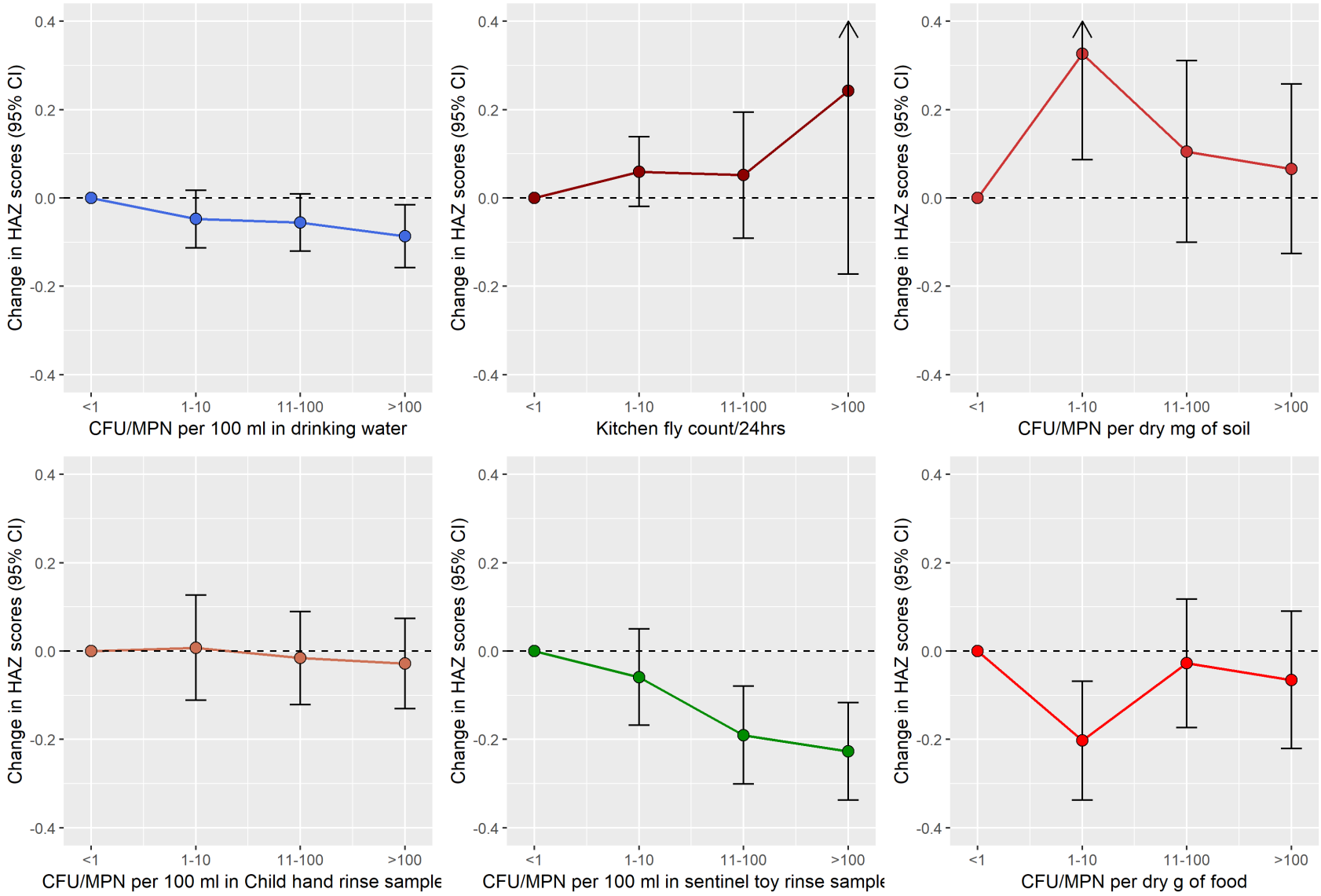


Figure B-10: Difference in HAZ score for a 1-log$_{10}$ median higher fecal contamination, stratified by log category compared to the reference level of no measured contamination

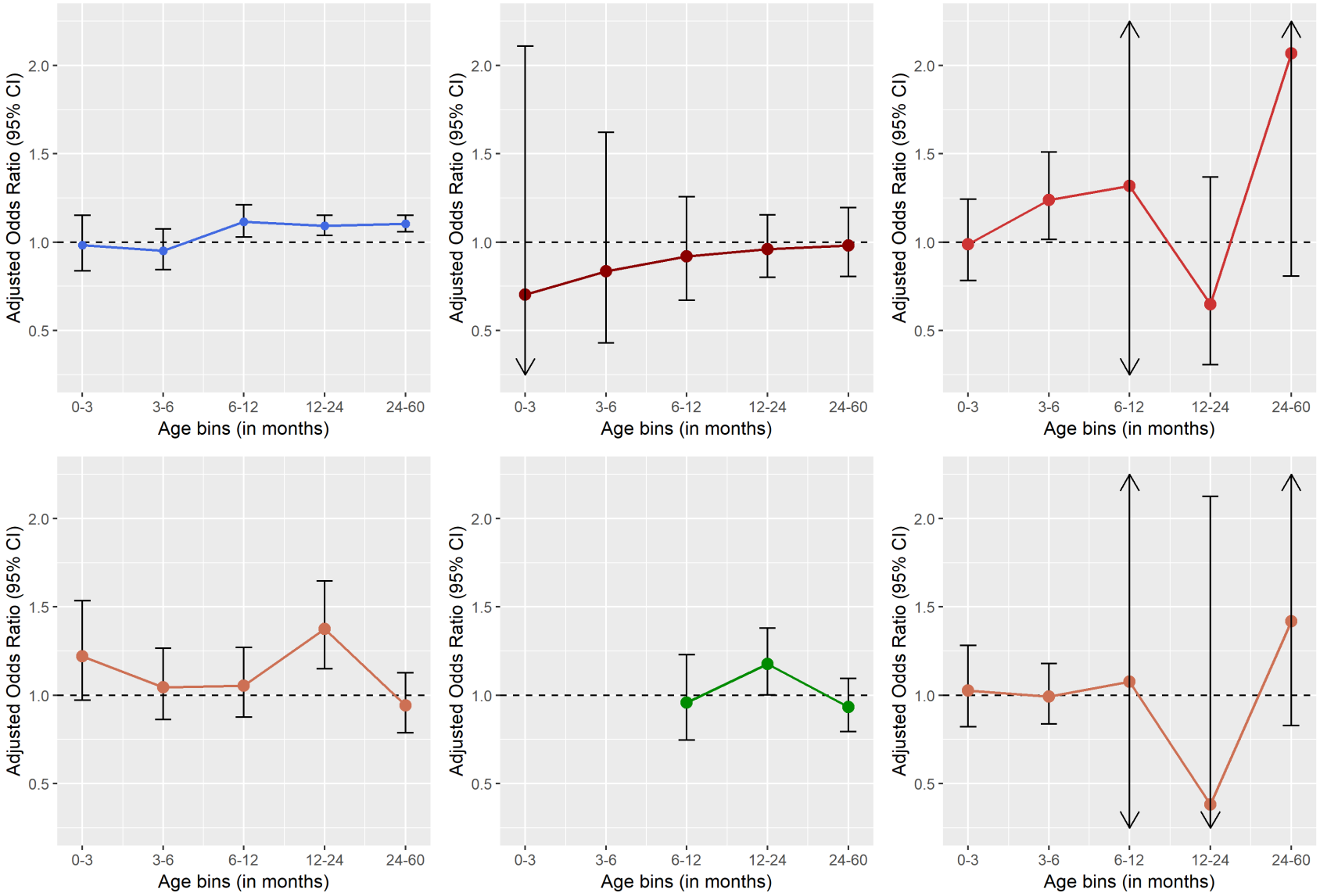


Figure B-11: Odds of diarrhea for 1-log_{10} higher fecal contamination, stratified by age

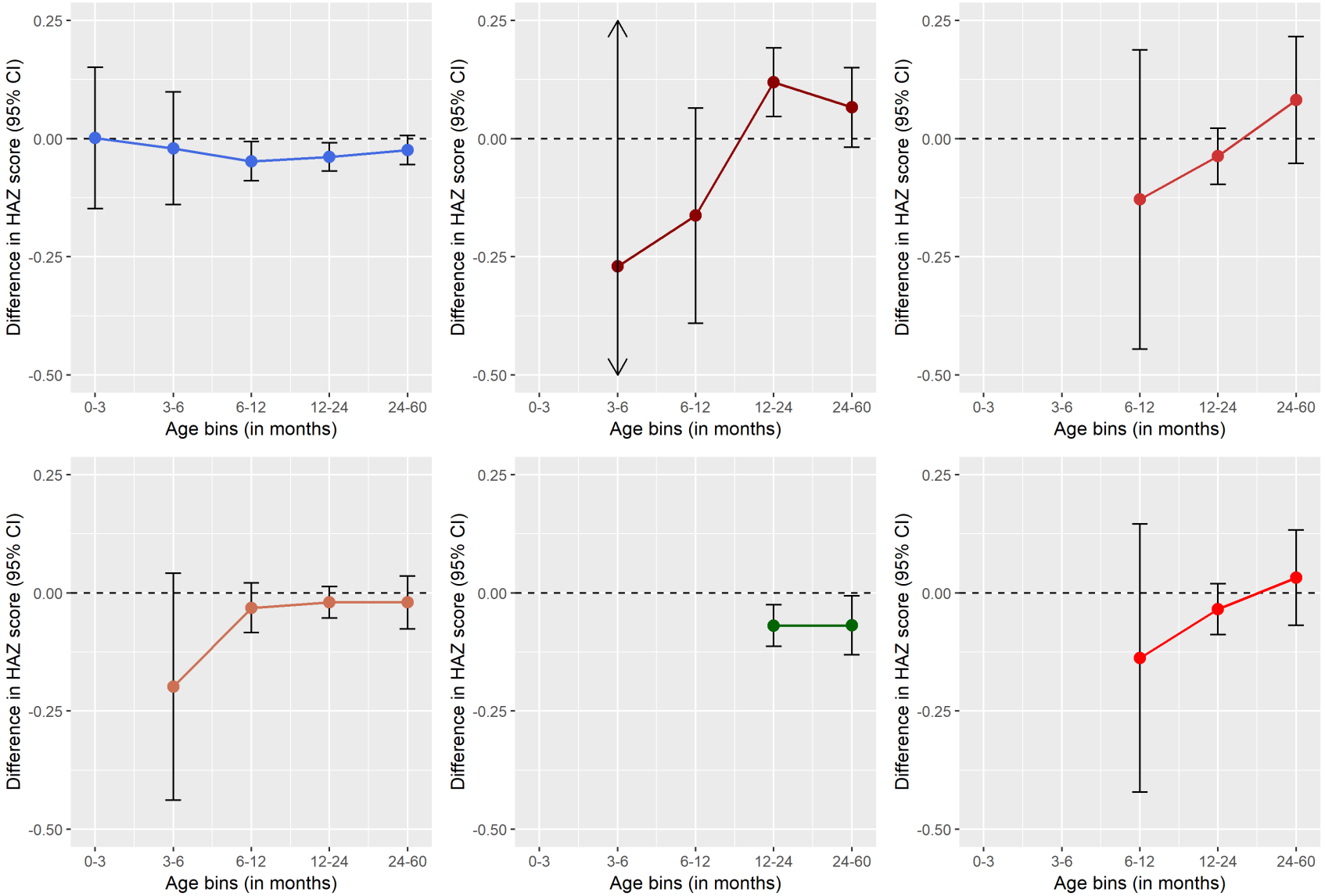


Figure B-12: Difference in HAZ score for a 1-log_{10} higher median in fecal contamination, stratified by age

7.3 **Appendix C -** Research Aim 1: Risk of Bias Assessments

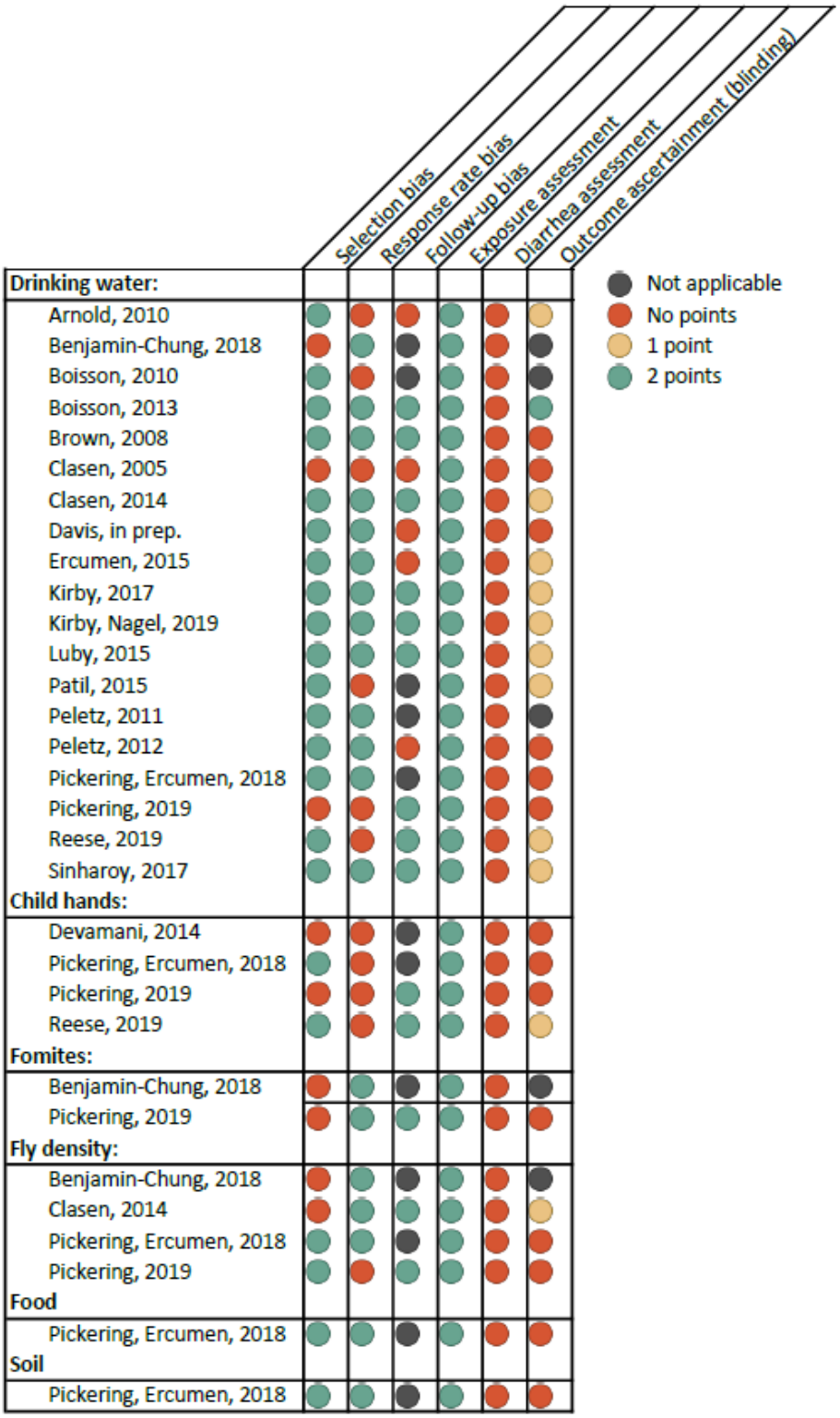


Figure C-1: Results from the risk of bias assessments for the diarrhea analyses

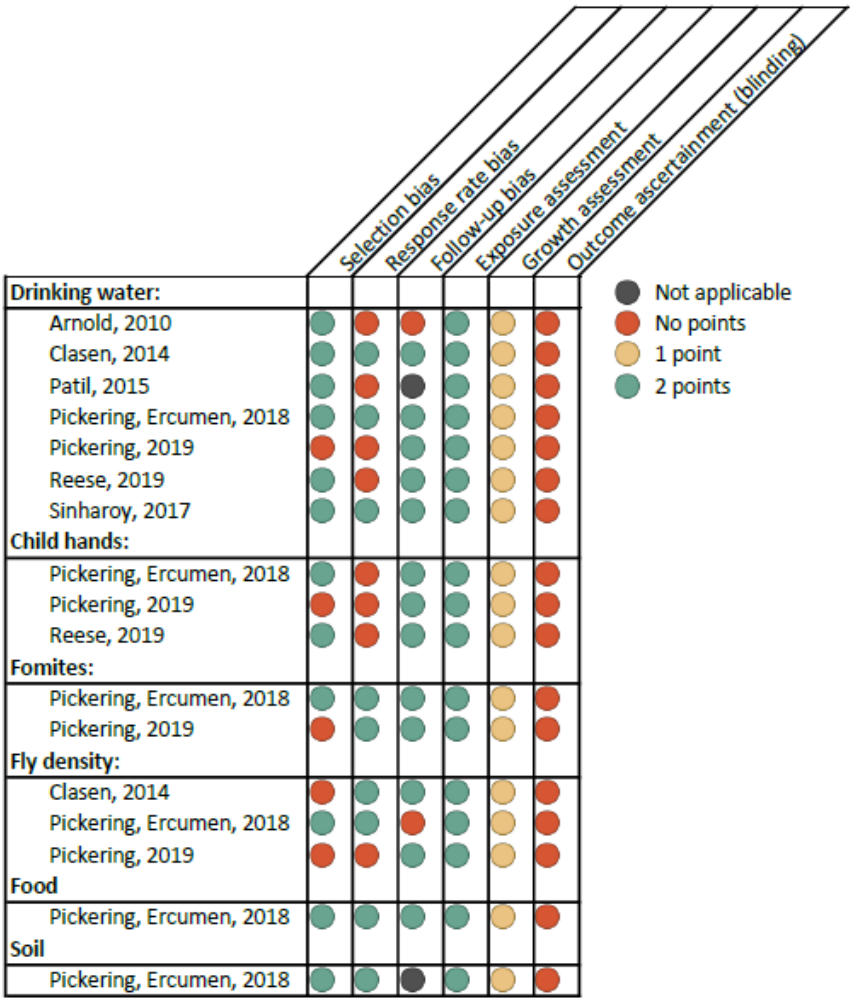


Figure C-2: Results from the risk of bias assessments for the linear growth analyses

7.4 **Appendix D** – Research Aim 2: Sensitivity Analyses

Household versus Community Exposure

The sensitivity analyses suggested that the attenuation estimates remained stables with both higher and lower assumptions for the odds of diarrhea for 1-$\log_{10}$ higher FIB concentrations in drinking water (Table D-1). They also indicated stronger attenuation when the maximum or minimum community water FIB concentrations were selected as the exposure, instead of the median (Table D-2).

Table D-1: Estimated attenuation factors for the odds of diarrhea for 1-$\log_{10}$ higher fecal indicator bacteria concentrations in drinking water when only household water fecal indicator bacteria concentrations z_t^* was used as the exposure variable for different household-community exposure scenarios, comparing higher and lower assumed odd to those used in the model (in bold).

	Household water Zt* / community water fecal contamination Wt assignments										
OR	100/0	90/10	80/20	70/30	60/40	50/50	40/60	30/70	20/80	10/90	0/100
2.0	1	0.69	0.62	0.57	0.53	0.49	0.44	0.41	0.39	0.32	0.13
1.5	**1**	**0.69**	**0.61**	**0.57**	**0.52**	**0.51**	**0.45**	**0.42**	**0.39**	**0.34**	**0.14**
1.1	1.02	0.71	0.68	0.55	0.56	0.52	0.45	0.41	0.41	0.36	0.17

Table D-2: Estimated attenuation factors for the odds of diarrhea for 1-$\log_{10}$ higher fecal indicator bacteria concentrations in drinking water when only household water fecal indicator bacteria concentrations z_t^* was used as the exposure variable for different household-community exposure scenarios, comparing the use maximum and minimum community water fecal indicator bacteria concentration to the median used in the model (in bold).

	Household water Zt* / community water fecal contamination Wt assignments										
Community WQ	100/0	90/10	80/20	70/30	60/40	50/50	40/60	30/70	20/80	10/90	0/100
Max	0.99	0.35	0.25	0.23	0.20	0.18	0.16	0.14	0.13	0.11	0.07
Median	**1**	**0.69**	**0.61**	**0.57**	**0.52**	**0.51**	**0.45**	**0.42**	**0.39**	**0.34**	**0.14**
Min	1.01	0.44	0.39	0.36	0.34	0.32	0.28	0.26	0.22	0.14	0.02

Single versus Multiple Samples

Similarly to the household versus community exposure analysis, the sensitivity analyses for the single versus multiple samples suggested that the attenuation estimates remained stable with both higher and lower assumptions for the difference in HAZ scores for 1-$\log_{10}$ higher FIB concentrations in drinking water (Table D-3). They also indicated stronger attenuation when the maximum household water FIB concentrations was used for multiple samples instead of the median, with more similar attenuation factors when the minimum was used (Table D-4).

Table D-3: Attenuation factors for the estimated difference in HAZ scores for 1-log_{10} higher fecal indicator bacteria concentrations in drinking water when only a single water sample z_t was used as the exposure variable for four scenarios if 1, 2 3 or 4 samples represented the household water fecal contamination z_t^*, comparing higher and lower assumed differences in HAZ scores to those used in the model (in bold).

	Number of samples representing household water fecal contamination Zt*			
HAZ	1	2	3	4
-0.4	1	0.56	0.52	0.45
-0.2	**1**	**0.56**	**0.52**	**0.43**
-0.1	1	0.56	0.53	0.44

Table D-4: Attenuation factors for the estimated difference in HAZ scores for 1-log_{10} higher fecal indicator bacteria concentrations in drinking water when only a single water sample z_t was used as the exposure variable for four scenarios if 1, 2 3 or 4 samples represented the household water fecal contamination z_t^*, comparing the use maximum and minimum household water fecal indicator bacteria concentration from repeat samples to the median used in the model (in bold).

	Number of samples representing household water fecal contamination Zt*			
WQ	1	2	3	4
Max	1.01	0.52	0.33	0.30
Median	**1**	**0.56**	**0.52**	**0.43**
Min	1	0.61	0.40	0.39