Rotavirus Transmission in Rural, Coastal Ecuador

by

Alicia N. M. Kraay

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in The University of Michigan
2018

Doctoral Committee:

Professor Joseph N. S. Eisenberg, Chair
Assistant Professor Marisa C. Eisenberg
Professor Edward L. Ionides
Professor Mark Wilson

ORCiD ID: orcid.org/0000-0002-7042-7048
uniqname: amullis

Last summer, one of my coworkers was collecting data for her dissertation in Ecuador in one of our study communities. While she was there, she watched a young child die a sad and gruesome death that could have been prevented with better access to medical care. When I lived in Ecuador for six months, I too saw many more children suffer on a daily basis from a lack of education and medical care. When I sit in my comfortable office, it is easy for me to think in abstract terms and forget that these stories of suffering are the daily experience of many children around the world. This dissertation is for these children and for the millions of other children like them. While I have no delusions that my work here has profoundly changed the world, I hope that the work I have done and continue to do can be a part of creating a new narrative for them–one that is more filled with hope.

ACKNOWLEDGEMENTS

I would first like to acknowledge my advisor, Joseph Eisenberg, for his support and mentorship over the past five and a half years–first, while I completed my M.P.H. and then as I continued on for my Ph.D. I am thankful for the amount of intellectual freedom he allowed me to have as well as for the opportunities that Joe encouraged me to pursue, even when they were not directly relevant to my dissertation.

I would also like to acknowledge several other faculty members. Thanks to Marisa Eisenberg and Mark Wilson, who both showed constant enthusiasm for their work, no matter how busy or overwhelming it got. Also thanks to Marisa for providing mathematical suggestions and career advice. Thanks to Mark Wilson, for providing useful feedback. Thanks to Ed Ionides, for making me think critically about statistics and distributions and for being patient as I struggled to learn how to use PoMP. Thanks to Betsy Foxman, who intentionally provided many professional development opportunities for me to learn more about biology and for career advice, despite not being a member of my committee. Thanks to Andrew Brouwer, for all of his mathematical expertise, generous coding help, and for teaching me so much about threshold analysis. Thanks to Gwenyth Lee, for coordinating data collection for chapter 4 and for useful discussions about planned analyses. Thanks to Hal Morgenstern, for allowing me to GSI in his course and for being patient with all of my detailed questions. I am also thankful for the field staff and study participants in Ecuador, for enrolling in the study and helping with data collection. Thanks to the Epidemiology department and the University of Michigan for encouraging transdisciplinary collaborations

and for providing so many opportunities to learn from incredible researchers all over the world.

In addition to my support from my committee and other faculty members, I have been blessed with a strong support network in Ann Arbor. My family and friends have provided so much support and encouragement along the way. My parents taught me from a young age to ask hard questions and always believed that I could accomplish anything I aimed to achieve. My sister has shown me so many times what it means to remain positive and to persevere through adversity. My co-workers in the EcoDess study and in the MIDAS lab group have provided good research advice as well as personal support and laughter. My friends from church have also helped me to maintain my perspective through difficult times and to take breaks and enjoy the journey.

I thank my husband, Bruce Kraay, for being a source of strength, love, and joy throughout my entire time in graduate school. Our early morning runs and cooking adventures kept me sane and balanced and helped me to enjoy the journey. Bruce is a light and a source of joy to all who know him, but especially to me. I am incredibly grateful for the life we have built together.

Finally, I thank God, for clearly guiding me to the field of epidemiology in general and to the University of Michigan specifically during the first few months that I lived in Ecuador. I am thankful that, through my travels, I was taught to see why this work matters and how it is a part of bringing His kingdom to the world. I have no doubt that this work is exactly what He has called and equipped me to do in this time.

This work was funded by various grants and scholarships: the National Institutes of Health (grant R01-AI050038), NIH MIDAS (grant U01GM110712), University of Michigan Regents Fellowship, Rackham Graduate Student Research Funds, and an Interdisciplinary Training in Microbial Systems (ITiMS) grant. Both the Epidemiology department and Rackham also provided generous financial support that allowed me to attend numerous

conferences during my time as a student.

PREFACE

A version of chapter 2 has been published in *Epidemiology*. A version of chapter 3 is under review at *Procedings of the National Academy of Sciences*. Chapter 4 is currently being prepared for submission. Various co-authors have contributed substantially to the development of each of the dissertation chapters. For chapter 2, Andrew Brouwer, James Trostle, William Cevallos, and Joseph Eisenberg are all co-authors. For chapter 3, Nan Lin, Andrew Brouwer, Philip Collender, Justin Remais, and Joseph Eisenberg are all co-authors. For chapter 4, Edward Ionides, Gwenyth Lee, William Cevallos, and Joseph Eisenberg are all co-authors.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Rotavirus is a major cause of under-five mortality, particularly in the developing world, yet how it is transmitted is not well understood. How and why rotavirus spreads through populations is a critical concern for both preventing disease and reducing the burden. In this dissertation, I consider how rotavirus spreads through a region by focusing on two important processes: human travel and hydrological transport. I also examine how the introduction of the rotavirus vaccine in a remote region of Ecuador has changed rotavirus transmission, affecting both direct and indirect pathways. I use a combination of regression analysis and simulation modeling to address these questions. In chapter 2, I consider the effect of human travel on rotavirus transmission. To do so, I use longitudinal data from 15 villages in rural, coastal Ecuador to identify important determinants of travel patterns and the extent to which these variables are stable over time. I then incorporate these predictors into a regional transmission model to assess how demographic heterogeneity in travel impacts regional risk of rotavirus and the implications of this heterogeneity for disease interventions. In chapter 3, I investigate the conditions in which rotavirus can be transmitted through water sources and how this pathway depends on both temperature and local hydrologic conditions. I use previously published empirical studies of temperature and rotavirus persistence in water sources to conduct a meta-analysis relating temperature to rotavirus decay in water sources. Then, I combine the resulting temperature-decay function with data on hydrological characteristics of our Ecuador study site to build a transmission model that accounts for both direct transmission and water-mediated transmission. Using this model, I assess the importance of waterborne transmission to amplification of rotavirus infection within communities and spread of rotavirus through hydrological trans-

port within a watershed. In chapter 4, I consider how the Rotarix vaccine has changed rotavirus transmission patterns in our study region. I combine data on rotavirus vaccination from local health posts with 10 years of data from a population based case control study and 18 months of diarrheal disease surveillance to estimate: (1) the *direct effect* of vaccination on the rate all-cause diarrhea, (2) the *overall effect* of vaccination on rotavirus infection and all-cause diarrhea both by age group and at a population level, and (3) the fraction of cases attributable to rotavirus in our study region.

CHAPTER I

Introduction

Diarrheal disease is one of the top causes of child mortality worldwide and rotavirus is the top cause of severe diarrhea for children under two [4, 5]. While public health efforts have been successful at reducing, particularly among young children [6–8]. In 2013 alone, the global burden of disease study estimated that 122,400 deaths in children under five worldwide were attributable to rotavirus [5]. While rotavirus infection is ubiquitous worldwide, most mortality occurs in Africa and Southeast Asia [9]. The introduction of two rotavirus vaccines worldwide appear to have contributed to this decline, with rotavirus explaining a smaller proportion of diarrheal cases in countries that have introduced vaccination [10]. In this introduction, I will review the biological features of rotavirus, its transmission routes, mechanisms of circulation through regions, and the impact of vaccination.

1.1 Biology of rotavirus

Rotavirus is a double stranded RNA virus with high genetic diversity. There are seven types of rotavirus (A-G), but rotavirus A is the predominant type that infects humans. Type A rotaviruses are classified further into serotypes on the basis of mutations in the G and P protein, which are the surface proteins targeted by the immune system [11, 12]. As of 2004, there were 14 different G protein serotypes and 14 P protein serotypes. In many

high income countries, four combinations of these G and P proteins (P[8]G1, P[4]G2, P[8]G3, and P[8]G4) were responsible for over 90% of all reported rotavirus infections, but in other low and middle income countries genetic diversity was much higher [12]. For example, these same four genotypes explained only 50% of all circulating strains in Africa and 68% in South Africa and Asia, where G9 was also an important serotype. This higher level of diversity may be partially attributable to spillover events between animals and humans [13]. For example, the G5 genotype has historically been found only in livestock, but in the early 2000s appeared to be emerging as an important serotype in South America [12].

The predominant circulating rotavirus strain varies both by transmission season and within seasons in the same country [12]. In one city in Australia G9 explained roughly 12% of rotavirus cases in another city it explained 80% of cases during the same season [12]. Various studies have shown changes in the predominant strains of rotavirus over time, and there is some evidence that vaccine introduction may have led to a relative increase in the prevalence of non-vaccine strains, particularly the G2 serotype [14, 15]. However, because the dominant rotavirus strain has shifted over time prior to vaccine introduction, these findings are not conclusive [12, 16].

The combination of high genetic diversity and propensity to re-infection make rotavirus especially hard to control. In general, natural infection with one strain of rotavirus confers some immunity against re-infection, but this immunity is generally strain-specific and also wanes over time [17]. For this reason, individuals who have previously had rotavirus can be re-infected with a different strain, although symptoms are generally less severe in secondary infections (e.g., rotavirus is an incompletely immunizing infection) [18]. Because the majority of severe rotavirus infections occur early in life with nearly all children experiencing their first infection before two years of age, rotavirus is still considered a childhood disease. However, adults can become infected with rotavirus and may also have symptoms

[19]. Relative to other enteric viruses, rotavirus has one of the lowest infectious doses and highest shedding rates [20, 21].

Rotavirus infects intestinal cells, where it destroys absorptive enterocytes and causes secretory diarrhea. Infected individuals may shed up to 10^{10} virus particles per gram of stool, whereas only 10 particles are needed to cause disease [11]. In general, symptoms resolve in about five days, but shedding may continue for up to three weeks [22]. After shedding, rotavirus can persistent on both contaminated surfaces (i.e., fomites) and in water sources for months. For fomites, persistence on surfaces is highly dependent on humidity [23–27] whereas rotavirus persistance in water sources is more strongly driven by temperature [28–36]. The implications of this persistence is explored more in depth in chapter 3 of this thesis. Rotavirus infection also interacts synergistically with other co-infecting pathogens, such that individuals that are infected with other pathogens are more likely to have symptoms if they are also co-infected with rotavirus [37].

1.2 Transmission routes

Rotavirus is transmitted via the fecal-oral route. Rotavirus is highly infectious and is also highly persistent both on fomites and in water sources. Because of the high shedding of rotavirus relative to its infectious dose, rotavirus tends to produce explosive outbreaks. For this reason, transmission through indirect pathways, which is inherently slower, is often not considered for rotavirus. The primary mechanism of spread for rotavirus is thought to be direct contact, with fomites (e.g., contaminated surfaces) as an intermediate [3, 38–41].

However, the combination of the generally seasonal pattern of rotavirus infection and high persistance of rotaviruses in water sources suggests that waterborne transmission might also play a role. In chapter 3, we explicitly explore this possibility. In support of this

possibility, some water-borne outbreaks of rotavirus have been reported [42]. In these outbreaks, attack rates were higher among adults than children, unlike typical transmission patterns. In general, rotavirus incidence is inversely associated with temperature, with higher rotavirus incidence being observed at lower temperatures [43–45]. In temperate climates, this association is seasonal in nature, and rotavirus occurs more commonly during the winter months [45]. In tropical climates, rotavirus occurs throughout the year. However, in these settings, rotavirus is still associated with transient downward shifts in temperature [43]. While some seasonality in rotavirus transmission might be explained by temperature-dependent changes in rotavirus persistence on fomites, which would impact direct transmission, the empirical associations between temperature and decay on surfaces is not adequate to explain this seasonality [23–27]. Instead, seasonal changes in fomite-mediated transmission are more likely to be driven by seasonal shifts in humidity, where drier conditions promote increased fomite-mediated transmission [43]. More work is needed to understand the factors underlying the seasonal nature of rotavirus infection and whether or not indirect transmission through water sources might explain some of this association.

1.3 Rotavirus circulation in regions with developing infrastructure

Human movement is an important determinant of the spatial spread of disease. For example, the spread of outbreaks of SARS [46], pandemic influenza [47, 48], and cholera [49] have all been tied to human travel. To capture how the spread of disease is affected by movement, mathematical models that include more than one community account for human travel either using data (cell phone data or survey data) [50] or simple, population-level approximations based on population size and distance [51–53]. These population-level proxies are frequently used when more detailed data are not available, particularly early in disease outbreaks. However, given the speed of urbanization, it is unclear if these

approximations will continue to be valid as travel infrastructure changes. For example, if remote communities are targeted for road access by construction projects, travel patterns might be more similar between communities than these population-level proxies would predict. Conversely, if highly isolated comunities are especially slow to acquire road access, these models might underestimate disparities in travel. Research that analyzes how travel patterns evolve in tandem with developing infrastructure would help reveal whether or not this concern is warranted. In chapter 2, we explicitly investigate this possibility by assessing if the relative frequency of travel over time has changed in proportion to baseline remoteness in a region with developing infrastructure.

While understanding the average travel behavior is important, more information about which demographic subgroups are more likely to travel could inform intervention design. In general, both survey data and cell phone data can provide information about relative travel frequencies between community levels. However, cell phone data generally does not provide detailed information about the demographics of travelers whereas survey data can be used to gather this information. In contrast, while unbiased, survey data generally underestimates the total volume of travel [50]. Understanding these demographic determinants of travel can be used in conjunction with other methods to model the spatial spread of epidemics. In chapter 2, we explore the relevance of demographic heterogeneity in travel to rotavirus transmission dynamics.

1.4 Vaccination

Two vaccines have been developed for rotavirus, and are now in widespread use worldwide. The Rotarix vaccine is the version most commonly used in low resources settings [54]. This vaccine requires two doses and is a monovalent, live, attenuated vaccine. Rotarix has been effective at reducing severe rotavirus worldwide, but the strength of this protection varies by context and depends both on compliance and the immune response

of immunized children. In particular, randomized controlled trials showed that Rotarix is more effective in developed countries than in the low resource settings where the vaccine is most urgently needed [55–58]. Overall estimates of efficacy range from 96% in Asia [56] to 61% in Africa [58]. There are many reasons for these differences, but at least some of the disparity is probably related to the immune response to the vaccine and the genetic diversity of circulating rotavirus strains[12, 59].

In remote regions, logistical difficulties with maintaining the cold chain and transportation to health posts to receive vaccines make achieving the maximal public health benefit difficult [60–62]. Additionally, even with proposed price cuts for GAVI-eligible countries, Rotarix vaccination is still far more expensive than most other vaccines included in the expanded program on immunization [63]. Because the rotavirus vaccine is imperfect, this high price tag has led some countries to conclude that vaccination is not cost-effective [64]. More studies are needed to understand the effectiveness of rotavirus vaccination in remote regions where barriers to access are high. In chapter 4 of this thesis, we assess the extent to which rotavirus vaccination has improved health outcomes in a relatively remote region of Ecuador.

In addition to direct benefits for vaccinated individuals, several studies have also shown benefits to older children who could not have been vaccinated, suggesting indirect effects [16, 65–67]. Due to the incompletely immunizing nature of rotavirus infection, these indirect effects on population risk were not expected [16]. Because of logistical difficulties with achieving full vaccination in low resource settings, indirect protection may be especially beneficial in these contexts and could enhance the overall cost effectiveness of vaccination for a given level of coverage [16]. In this context, establishing the benefits of rotavirus vaccination in low to middle income countries and how these benefits relate to herd protection is a key research need. We explore the impact of Rotarix vaccination on these older, unvaccinated populations as well as younger children in chapter 4 of this

thesis.

1.5 Study region

The data used in this dissertation was collected as part of a larger study in rural, northwestern coastal Ecuador, where there has been dramatic environmental and social change due to the construction of a new road [68]. Prior to the start of the study, most communities in the region lacked road access. In 2003, a road construction project was completed that connected the region with the rest of the country. During the course of the study, new roads continued to be built, increasing road access for previously roadless communities. The study was conceptualized as a natural experiment to document changes that occurred in the region over time as a result of this increased road access.

The region covered by the study comprises approximately 150 villages located along three rivers (the Santiago, Onzole, and Cayapas rivers). These three rivers have distinct hydrological characteristics–the Onzole river is slowest and has both the highest turbidity and warmest temperature. The Santiago and Cayapas rivers have much lower turbidity and the Santiago is the fastest flowing river [69]. The region has a tropical climate and the water temperature for all three rivers is over 24° C. Most study communities lack piped water and rely on river water to meet their needs. Access to sanitation facilities is highly heterogeneous throughout the region, with some households having flush toilets and others engaging in open defecation.

The urban center of the region is Borbón, a city of about 5,000 people. Most of the population of this region is Afro-Ecuadorian, with a few Chachi communities and a growing Mestizo population. Communities that were enrolled in the study were selected to be a representative sample of the area overall and varied in their remoteness from Borbón. In total, 31 villages were followed at some point during the ten year study. During the course

of the study, some communities were dropped and others were added. Fifteen communities were followed for the full ten year period.

1.6 Dissertation overview

In this dissertation, I aim to contribute to a greater understanding of how rotavirus is transmitted, particularly in rural settings. In chapter 2, I explore the role of human travel in the spread of rotavirus throughout our study region and investigate how accounting for heterogeneity in travel leads to additional insights above and beyond modeling the average behavior. In chapter 3, I assess the potential role of water-mediated transmission of rotavirus and use the result to help partially explain rotavirus seasonality and how transmission pathways might vary by context. In chapter 4, I investigate how the introduction of the Rotarix vaccine has altered rotavirus transmission in rural, coastal Ecuador, suppressing both direct and indirect transmission pathways.

CHAPTER II

Determinants of Short Term Movement in a Developing Region and Implications for Disease Transmission

2.1 Introduction

Human mobility contributes to the spread of infectious disease at multiple spatial and temporal levels [70–74]. Human migration–the long term process by which individuals permanently or semipermanently relocate–has been implicated in the introduction of pathogens to new locations. For example, movement along the Silk Road introduced the Black Death to Europe in the 14th C., and the forced transatlantic migration of African slaves brought malaria to the US [70]. Short-term movement–travel to another location with subsequent return–can also introduce pathogens when it occurs over large distances [71]. Examples of this process include the spread of SARS [46], pandemic influenza [47, 48], and cholera [49]. Short-term movement also occurs over smaller distances more critical for pathogen circulation, analysis of which can help reveal the spatial patterns of infection spread. Short-term movement has been implicated in the 2010 cholera epidemic in Haiti [75] as well as seasonal influenza [76] and dengue transmission [74, 77].

Road availability can influence both frequency and determinants of mobility. For example, as remote towns gain road access, travel frequency and average distance may increase,

and towns' sociodemographic environment may change [78]. Easier access to roads also makes urban markets more accessible to rural farmers, which may influence their economic activities within the village as well as reasons for their travel [79]. Finally, road access may also facilitate migrant labor, leading to a more transient population and changing social network structure [68, 79, 80].

Therefore, in addition to the benefits of road access, construction of new roads might also lead to increased regional risk of disease [68]. Other work has shown that environmental change occurring in tandem with road construction projects may lead to increased local risk of malaria and dengue [81, 82]. As previously remote regions gain access to new roads in parts of Asia, Africa, and Latin America, their changing travel patterns may facilitate both pathogen introduction and pathogen circulation, leading to convergence of regional risk.

The relevance of both short- and long-term movement for transmission has been widely recognized by transmission modelers, who regularly incorporate movement into their models [51, 75, 76, 83, 84]. For example, travel can increase transmission and allow for pathogen invasion even when there would not otherwise be an outbreak [85–87]. Most of this work has focused on developing models that adequately capture the average rate of travel and distribution of destinations rather than the variability in these patterns by demographic subgroups [51–53]. A separate body of work has looked at the role of population heterogeneity in disease transmission and has shown that outbreaks can occur when the average basic reproduction number ($\mathcal{R}_0$) is less than one if a subgroup of superspreaders is at higher risk [88, 89]. We aim to combine these two lines of research, investigating how population heterogeneity in both disease susceptibility and travel frequency may influence outbreaks. We use a transmission model, parameterized to reflect rotavirus, as a case study because it is epidemiologically important in our study region and globally and because its attack rate differs by age [3, 4, 37].

This paper contributes to a more developed understanding of short-term movement. Using data from a region in rural Ecuador that has been experiencing continuous road development, we evaluate the changes in movement patterns and village demographics over a ten year period (2003-2013) following the construction of a new road. We next identify determinants of short-term travel outside the study region and evaluate the stability of these determinants over time. Finally, we present a transmission model for rotavirus in our study region that accounts for heterogeneity in both travel and disease transmission and illustrates how this heterogeneity can lead to both increased risk of disease and changes in disease etiology.

2.2 Methods

2.2.1 Data

Data for this project were collected in rural Ecuador in conjunction with a larger study [68]. In 1996, the government of Ecuador began a road construction project to link the Andes with the Ecuadorian Pacific coast. The highway was completed in 2001, but new roads continue to be built, increasing access for previously more remote villages. Our EcoDess project was initiated in 2003 as a natural experiment to document longitudinal changes that occurred in villages surrounding the local metropolitan center of Borbn due to road construction. The present analysis focused on 15 study villages that were followed from 2003-2013. See reference [68] for more details about the study region.

In this paper, we use data from movement surveys conducted in 2003, 2007, 2010, and 2013 and census data collected by our study at about the same times. We use stool sample positivity for rotavirus from our ongoing case-control study from 2007–before the vaccine was introduced [90]–to parameterize our transmission model [91]. The methods used to identify rotavirus infection are described in detail elsewhere [37]. Our analyses considered

15 villages, 1,347 households, 5,443 individuals, and 1-4 time points per individual, for a total of 10,725 observations and a median of 2 observations per individual.

2.2.2 Software

Time trend analyses were conducted in R (v. 3.1.2) using the package gee, and other multivariate analyses were conducted in SAS (v. 9.4). Transmission modeling was conducted in R (v. 3.1.2) using the package deSolve.

2.2.3 Defined measures

2.2.3.1 Remoteness

Remoteness reflects a number of interrelated factors that together capture the relative isolation of a given community. Here, we assess remoteness using two measures: 1) travel frequency, which varies over time, and 2) the feasibility of travel at baseline, based on cost and travel time to reach the nearest large town. For the duration of this paper we refer to the latter metric (cost and travel time) as 'baseline remoteness.' Travel frequency The main outcome of interest in this study was whether an individual had traveled outside of the study region in the past seven days. Baseline cost and travel time Project staff used data collected at the start of the study to develop a metric for the remoteness of each village based on the cost and travel time required to reach Borbón, the largest town in the region [68]. This measure was subsequently categorized into three groups: close, medium, and far. While this metric reflects the feasibility of travel, actual travel frequency may not necessarily be shaped only by road accessibility; thus, the two metrics may measure different aspects of remoteness. Even if the absolute cost and travel time to Borbón changes over time as roads develop, if the relative isolation of these communities remains constant, remoteness would remain an important predictor of travel. If, however, road

development homogenizes travel frequency across the region, any differences in travel by baseline remoteness should diminish over time. To determine whether or not the regional risk appeared to be converging, we considered baseline remoteness as a covariate in all multivariate models and investigated whether the association between baseline remoteness and travel frequency was stable over time.

2.2.3.2 Covariates

Occupation was classified as none, domestic, student, agriculture, salaried, or other. Age was calculated using date of birth and was categorized into three groups: $<$5, 5-13, and $>$13 years. Each individual's duration of residence in the village was analyzed using z-scores calculated within each age bracket. Highest household education was the highest number of years of schooling reported by any individual of that household. Village-level availability of secondary schools was included in the models as a binary variable and was treated as time-invariant. The mean duration of residence for each village for all individuals over the age of 13 was also included as a proxy for population stability. All of these variables were taken from our census data.

2.2.4 Regression models

To investigate time trends and identify determinants of travel, we used a generalized estimating equation (GEE) framework, which allowed us to account for community clustering but retain a population average interpretation [92]. Because GEE models are not based on maximum likelihood estimation, we compared model fit using quasi-information criteria (QIC) [93]. Change-over-time was first evaluated by building regression models containing only the covariate of interest and years since 2003. These time trends were conceptualized as providing descriptive information about the study region and are therefore unadjusted. To identify determinants of travel, forward regression was used. Because

all variables except baseline remoteness and secondary school availability were allowed to change over time, our model also adjusts for time-dependent confounding. The stability of associations over time was evaluated by comparing, for each study year, the model containing only cost and travel time, the model containing all variables except cost and travel time, and the full model.

2.2.5 Transmission model

We developed a disease transmission model with a susceptible, infectious, recovered (SIR) framework (parameterized to reflect rotavirus) to investigate the impact of travel to an urban center (here modeled as analogous to Borbón) on a small village. Because the geography and transportation network of the region causes most travel to traverse Borbón, almost all travel outside the region is to another urban setting, and travel to the village by city residents is assumed to be negligible, this relatively simple two-community model suffices to investigate the impact of travel by members of the village on their community as a whole. To capture heterogeneity in risk of infection and travel, we stratified the modeled populations into two age categories (under and over 5 years), with children under five being more likely to become infected and less likely to travel (see Fig. 2.1 and the appendix for additional details on model structure). The appendix also presents the equations, parameterization, a sensitivity analysis of the transmission parameters, and calculations of the basic reproductive number for each community in isolation, which we call $\mathcal{R}_0$*. To understand how disease risk (measured by cumulative incidence) might change with time, we modeled an average close, medium, and far village, using travel frequency and population size in the data for each of the four study years (2003, 2007, 2010, 2013) for a total of 12 models (3 villages and 4 time points). To examine the separate roles of demographic heterogeneity in travel and the level of travel, we conducted model analyses to investigate changes in cumulative disease incidence as i) we increase the average travel frequency for

all village residents (assuming children and adults travel the same) and ii) we fix the travel frequency of children but increase the travel frequency of adults. These analyses used the population size and structure from 2013 and considered rotavirus transmission rates corresponding to the maximum and minimum observed in the study region (corresponding to $\mathcal{R}_0$* of 0.79 and 1.44 respectively).

2.3 Results

2.3.1 Evaluating demographic change

Baseline characteristics of the participants are shown in Table 2.1. Over time, the fraction of individuals traveling outside the region increased for all remoteness levels, with a 9.1%, 8.2% and 4.4% increase in travel for close, medium, and far villages respectively (see Table 2.1 and Figure 2.2). In 2013, far villages had similar travel patterns to the close villages in 2003. Far villages consistently had lower levels of travel than the medium or close villages with an average 6.0% and 3.2% difference comparing far with close and medium villages respectively, suggesting that travel patterns for the communities had not yet converged during the study. All far villages that showed an increase in travel patterns did not begin their increases until 2010, suggesting a time lag in the effect of changing infrastructure for more distant communities.

In addition to increased travel, there was also a tendency toward increased socioeconomic status. The proportion of individuals having no occupation decreased over time. The fastest growing occupation was salaried worker, and this change was especially pronounced in the most remote villages. The proportion of individuals working in agriculture decreased slightly for close and far villages.

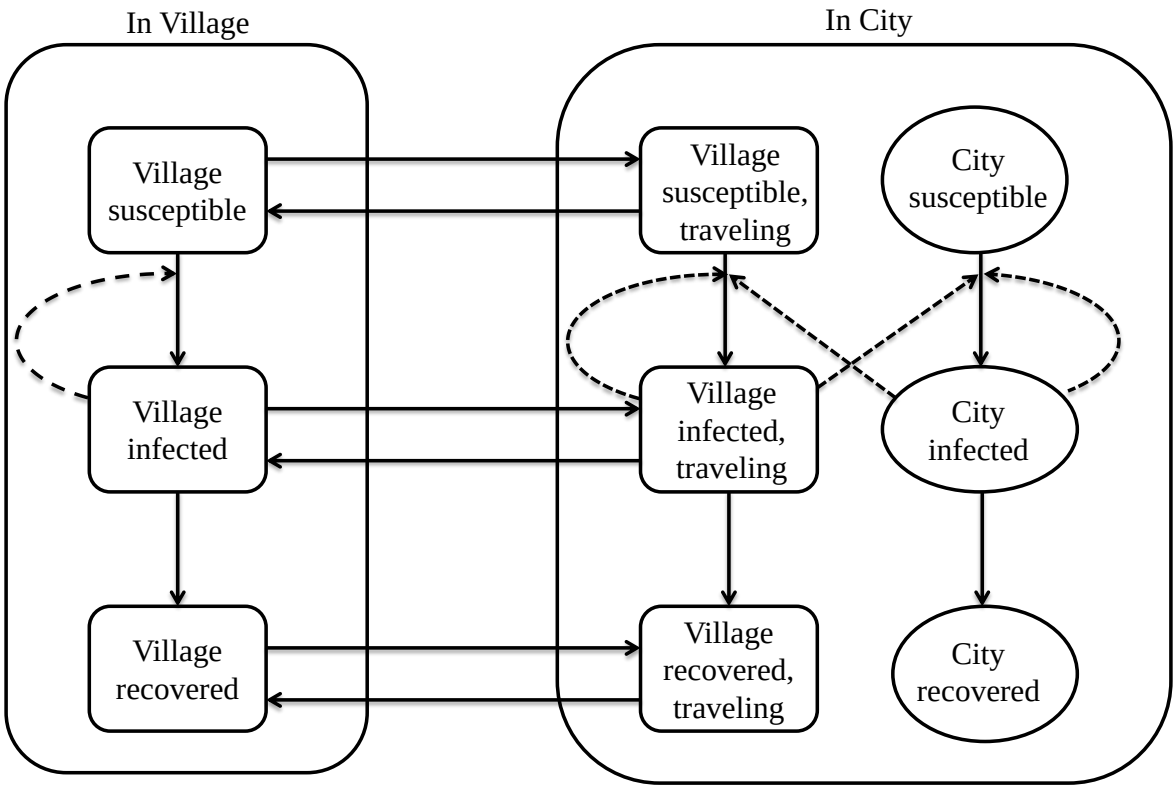

Figure 2.1: Rotavirus transmission model diagram. Dashed arrows represent transmission events and solid arrows represent movement of people. Villagers and city residents are classified as susceptible (S), infected (I) and Recovered (R), with separate compartments for individuals under age 5 and ages 5 and older (not shown). Susceptible people can become infected by direct transmission within their own community. Villagers can infect and be infected by the city during travel. We ignore travel by city dwellers and between smaller villages. Village SIR model is stratified by baseline remoteness (close, medium and far) and both village and city SIR models are stratified by year (2004, 2007, 2010, and 2013) resulting in 12 models. Each model is stratified by age (less than or greater than 5 yrs.). This stratification results in 8 transmission parameters (6 village-level due to 2 age-groups and three community groups, and 2 city level due to two age groups), and 24 travel rate parameters to and from the city (2 age groups, 4 years, and 3 community types, where). Population size also varies by age, community type, and year. The recovery rate parameter is the same for all SIR models. See supplement A.2 for model equations and details.

Table 2.1: Sample characteristics by stratum of remoteness. For baseline values, continuous variables are presented as mean (standard deviation) and categorical variables are presented as percent (n). Time trend values represent the correlation between time and the covariate accounting for community level clustering using a GEE with an independence correlation structure.

	Remoteness: Close (N=4,748†)		Remoteness: Medium (N=1,440†)		Remoteness: Far (N=4,534†)	
	Baseline (SD)	Time trend β (SE)	Baseline (SD)	Time trendβ (SE)	Baseline (SD)	Time trend β (SE)
Village variables						
Cost/travel time	0.018	—	0.07	—	0.17	—
Mean duration of residence	18.2 (6.5)	0.215 (0.096)	24.4 (3.0)	0.433 (0.200)	28.1 (3.4)	-0.037 (0.187)
Community age	23.3 (1.6)	0.085 (0.056)	25.2 (3.2)	0.434 (0.278)	23.6 (3.1)	0.223 (0.088)
Individual and household variables						
Highest household education	7.0 (3.8)	0.080 (0.051)	5.6 (3.6)	0.203 (0.026)	6.2 (3.2)	0.139 (0.038)
Housing index	5.01 (.93)	0.019 (0.007)	4.60 (0.94)	0.039 (0.031)	4.37 (1.20)	0.034 (0.031)
Duration of residence (z-score)	-0.18 (0.99)	0.001 (0.006)	0.01 (0.98)	-0.012 (0.012)	0.13 (0.93)	-0.008 (0.005)
Age (years)	37.9 (18.4)	0.155 (0.065)	41.8 (20.4)	0.826 (0.693)	38.2 (18.8)	0.111 (0.283)
Age						
<5	21.2 (216)	-0.029 (0.010)	21.3 (70)	-0.045 (0.085)	19.6 (195)	-0.055 (0.025)
5–13	25.7 (261)	0.003 (0.004)	22.6 (74)	-0.115 (0.057)	28.8 (287)	0.027 (0.019)
>13	53.1 (540)	0.015 (0.008)	56.1 (184)	0.124 (0.010)	51.6 (514)	0.011 (0.026)
Male	51.9 (523)	-0.005 (0.007)	50.8 (165)	-0.005 (0.009)	49.8 (491)	-0.003 (0.008)
Occupation						
None	22.2 (223)	-0.022 (0.008)	28.1 (88)	-0.058 (0.071)	24.8 (230)	-0.042 (0.018)
Domestic	16.8 (169)	0.012 (0.006)	18.5 (58)	0.039 (0.049)	18.8 (174)	0.003 (0.015)
Student	32.7 (328)	0.012 (0.012)	21.1 (66)	-0.031 (0.027)	31.4 (291)	0.053 (0.018)
Agriculture	19.8 (199)	-0.022 (0.011)	24.7 (80)	0.025 (0.040)	20.1 (186)	-0.064 (0.024)
Salaried	4.4 (44)	0.048 (0.014)	2.6 (8)	0.191 (0.040)	2.3 (21)	0.065 (0.034)
Other	4.1 (41)	0.019 (0.019)	2.9 (9)	0.004 (0.064)	2.7 (25)	0.055 (0.039)
Outside travel	8.2 (83)	0.051 (0.028)	7.1 (23)	0.189 (0.016)	3.9 (38)	0.090 (0.030)

†: Includes individuals with missing data on community of residence (n=3). These individuals are not included in Tables 2.3 or 2.4.

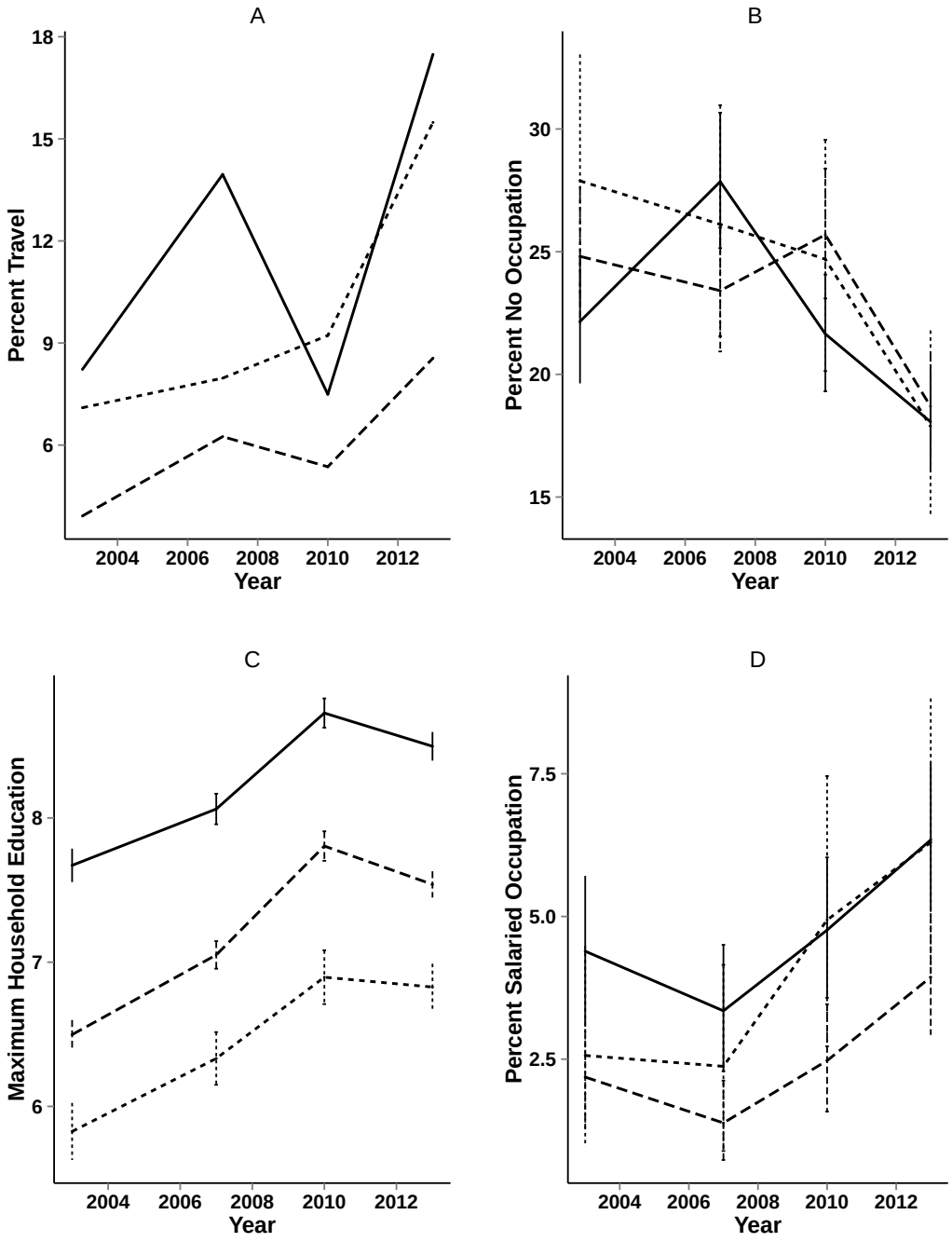

Figure 2.2: Demographic changes over time by study year and remoteness strata from the survey data. Close, medium, and far villages are shown as solid line, small dashed line, and longer dashed line respectively. A) Percent out-of-region travel (weighted by population size), B) Percent reporting no occupation, C) Maximum household education (years), D) Percent reporting salaried occupation.

2.3.2 Identifying predictors of travel

Gender was not associated with travel frequency and was therefore not included in the regression models. Based on prior knowledge, we conceptualized remoteness as being an 'exposure' whose effects were mediated by various demographic factors. However, to estimate the effect of these mediators on travel, we adjusted our demographic models for remoteness because it was a common cause of the various demographic characteristics and travel and therefore a confounder of those associations. See supplement A.1 for causal diagram.

Because the relationship between age and travel patterns was non-linear, we used a categorical age variable ($<$5, 5-13, $>$13). The results of model building are shown in Table 2.2. Although all variables except for age were associated with both remoteness and out-of-region travel (Table 2.2), only secondary school availability appeared to be markedly confounded by remoteness (comparing the Unadjusted and Remoteness Adjusted models). Secondary school availability and mean duration of residence appeared to be the strongest mediators of the remoteness-travel estimate, as indicated by the reduction in the point estimate (comparing the Unadjusted Model with Model 1). In general, occupation and education appear to have independent effects. However, the point estimate for salaried occupation attenuated after adjusting for education, most likely because of the much higher required education for salaried workers. In other words, there was no association between education and occupation except for salaried workers. The effect of occupation also appeared to be confounded by age (comparing Models 3 and 4).

After adjusting for all variables, travelers were more likely to come from households with higher education (OR=1.10, 95% CI 1.08, 1.12) and to work either in domestic (OR=1.64, 95% CI: 1.14, 2.36), agriculture (OR=1.50, 95% CI: 1.13, 1.98), or salaried work (OR=3.27, 95% CI: 2.15, 4.96). Remoteness remained a strong predictor of travel (OR=0.51, 95% CI: 0.38, 0.67), with travel being more common among individuals living

Table 2.2: Determinants of travel patterns. We used forward generalized estimating equation (GEE) regression with compound symmetry covariance structure, and all terms are odds ratios (95% CI). First, the unadjusted associations with each variable and the outcome were considered. To evaluate potential confounding by remoteness, each unadjusted association was adjusted for remoteness only. A series of models was used to investigate the relationships between the remaining variables. Model 1 contained only village-level variables. Model 2 was further adjusted for education. Model 3 further adjusted for occuapation. Model 4 included all variables in model 3 as well as duration of residence and age.

	Unadjusted	Remoteness Adjusted	Model 1 Village	Model 2 Model 1+Education	Model 3 Model 2+Occupation	Model 4 Fully Adjusted
Village variables						
Remoteness	0.69 (0.59, 0.82)	0.69 (0.59, 0.82)	0.48 (0.35, 0.66)	0.46 (0.34, 0.61)	0.51 (0.38, 0.68)	0.51 (0.38, 0.67)
Mean duration of residence (yr)	1.06 (1.02, 1.10)	1.06 (1.02, 1.10)	1.05 (1.01, 1.08)	1.05 (1.02, 1.08)	1.04 (1.01, 1.08)	1.04 (1.01, 1.07)
Secondary school availability (y/n)	1.47 (1.11, 1.95)	0.76 (0.58, 1.00)	0.61 (0.40, 0.94)	0.50 (0.35, 0.70)	0.56 (0.40, 0.78)	0.53 (0.37, 0.76)
Individual and household variables						
Highest household education (yr)	1.11 (1.09, 1.13)	1.10 (1.08, 1.13)	–	1.11 (1.09, 1.13)	1.10 (1.08, 1.13)	1.10 (1.08, 1.12)
Occupation						
None	1.00 (Ref)	1.00 (Ref)	–	–	1.00 (Ref)	1.00 (Ref)
Domestic vs. None	2.77 (1.99, 3.87)	2.70 (1.96, 3.73)	–	–	2.89 (2.10, 3.99)	1.64 (1.14, 2.36)
Agriculture vs. None	2.54 (1.81, 3.57)	2.47 (1.78, 3.43)	–	–	2.71 (1.97, 3.71)	1.50 (1.13, 1.98)
Salaried vs. None	7.82 (5.27, 11.58)	7.37 (4.91, 11.07)	–	–	5.69 (3.95, 8.20)	3.27 (2.15, 4.96)
Other vs. None	4.51 (2.89, 7.03)	4.36 (2.78, 6.84)	–	–	4.21 (2.57, 6.90)	2.43 (1.51, 3.92)
Duration of residence (yr)	0.99 (0.90,1.10)	1.00 (0.91, 1.09)	–	–	–	1.04 (0.95, 1.13)
Age						
<5	1.00 (Ref)	1.00 (Ref)	–	–	–	1.00 (Ref)
5–13	0.75 (0.59, 0.95)	0.77 (0.62, 0.95)	–	–	–	0.65 (0.50, 0.85)
>13	2.94 (2.25, 3.85)	2.83 (2.15, 3.73)	–	–	–	1.73 (1.30, 2.31)

in less remote regions. Individuals living in areas with a secondary school were also less likely to travel than people without secondary schools (OR=0.53, 95% CI: 0.37, 0.76). The effect of age was non-linear; individuals aged 5-13 were less likely to travel than children under five (OR=0.65, 95% CI: 0.50, 0.85) and adults over the age of 13 were more likely to travel than children under five (OR=1.73, 95% CI: 1.30, 2.31).

Although the effect estimate for increased travel by household level of education is small for each year increase in education (OR=1.10), the cumulative effect is substantial. For instance, a five-year difference in household education (roughly the difference between completing primary and secondary school) would correspond to an odds ratio of 1.61 for traveling.

Overall, the full model, including both baseline remoteness (cost and travel time) and all demographic variables was the best model for all four study years (Table 2.3). The QIC values including demographic factors were consistently lower than those with remoteness only, suggesting that these variables combined provided more information than the baseline remoteness metric alone. Although the cost and travel time metric provided less information than the demographic variables, adding it to the model improved model fit for each study year, supporting its continued use as a measure of remoteness in our region.

Table 2.3: Predicted cumulative incidence per 1,000 people by remoteness and study year. Models were parameterized to reflect data collected in our field site over time and from the literature regarding the three disease processes: travel to the city (by community, age group, and study year), transmission rate (by community and age group, but not time), and recovery rate (estimated from literature at five days). Community population size and the fraction of children under 5 years of age were also estimated for each study year using survey data and were incorporated into our model. The population of the city was fixed at 5,000 for all study years. All parameter values are shown in the supplement section A.3 along with relevant derivations.

Remoteness	2003	2007	2010	2013
Close Villages	13	24	13	28
Medium Villages	12	16	16	31
Far Villages	7	10	10	14

2.3.3 Travel and disease transmission

2.3.3.1 Component effects

To disaggregate the impact of travel from the impact of heterogeneity in travel by age, we present simulated cumulative disease incidence that consider change in i) the overall travel rate for all ages (homogeneous travel) and ii) a proportional increase in travel for adults only (heterogeneous travel by age).

Homogeneity in travel As travel increased from levels seen in far villages to those in close villages, the predicted disease incidence rate doubled (15, 27, and 31 per 1,000 people for close, medium and far villages respectively, where $\mathcal{R}_0$*=0.79 for all village types). While higher transmissibility within the village ($\mathcal{R}_0$*) increased risk, the effect of travel was much stronger. In general, predicted incidence at the maximum transmission rate ($\mathcal{R}_0$*=1.44) was only 1-2 cases per 1,000 people higher than risk at the minimum transmission rate ($\mathcal{R}_0$*=0.79), less than a 7% increase (details in supplement, section A.4).

Heterogeneity in travel As the relative travel rates of adults increased compared to children, the risk of infection for the village increased, suggesting that the adults disseminate the infection into local communities (see Table 2.4). Although this increase in risk was greatest for adults, risk also increased for children. Only results from close villages are shown; results were qualitatively similar for medium and far villages. Based on our survey data (Table 2.2), adults had 1.73 times the travel of children, so we would expect our study region to most closely resemble the medium travel heterogeneity scenario (scenarios defined in table 2.4).

Further, as heterogeneity in travel increased, the attributable fraction (AF) of infections acquired locally increased for children, but decreased for adults (Figure 2.3). The magnitude of local transmission depended on transmissibility within the village. Under conditions of high heterogeneity in travel and high transmission, 24% of all transmission to children

Table 2.4: Percent difference in cumulative incidence predicted by the transmission model for adults, children, and the whole community for increasing heterogeneity of travel by age group. Increasing heterogeneity was done using a proportionality constant where $travel_{adult} = ctravel_{child}$ and c takes values of 1 (low), 2 (medium), or 3 (high). Low-within village transmission corresponds to a baseline R_0 of 0.79 and high within village transmission corresponds to a baseline R_0 of 1.43 (R_0 increases slightly with increasing heterogeneity)

Travel heterogeneity	Adults	Children	Community
Low within-village transmission			
Medium vs. None	1.6%	0.2%	1.4%
High vs. None	3.1%	0.3%	2.8%
High within-village transmission			
Medium vs. None	1.7%	0.3%	1.4%
High vs. None	3.3%	0.6%	2.8%

originated within the village, compared to 6% for no heterogeneity in travel (i.e., adults traveled the same as children) and low transmission. However, regardless of whether $\mathcal{R}_0$* was greater than 1, most transmission for both age groups occurred in the city. This result held over the transmission parameter sensitivity analysis, but was modified by village population size. Increasing the population of the village relative to the city led to proportional increases in the AF for local transmission but slight decreases in the overall incidence (see supplement section A.6).

2.3.3.2 Predicted Incidence

Over time, modeled risk increased for all three types of communities (Table 2.5). In 2013, the risk in far villages was similar to that seen in close villages in 2003. At the end of the study, the predicted risks for close and medium villages were similar, but the risk for far villages remained lower. Thus, although all villages had predicted increases in risk over the study period, the gradient of risk by remoteness was preserved for the most remote villages. Comparison with our component effects model suggests that this increase in risk was almost entirely driven by increased travel rather than changing demographics.

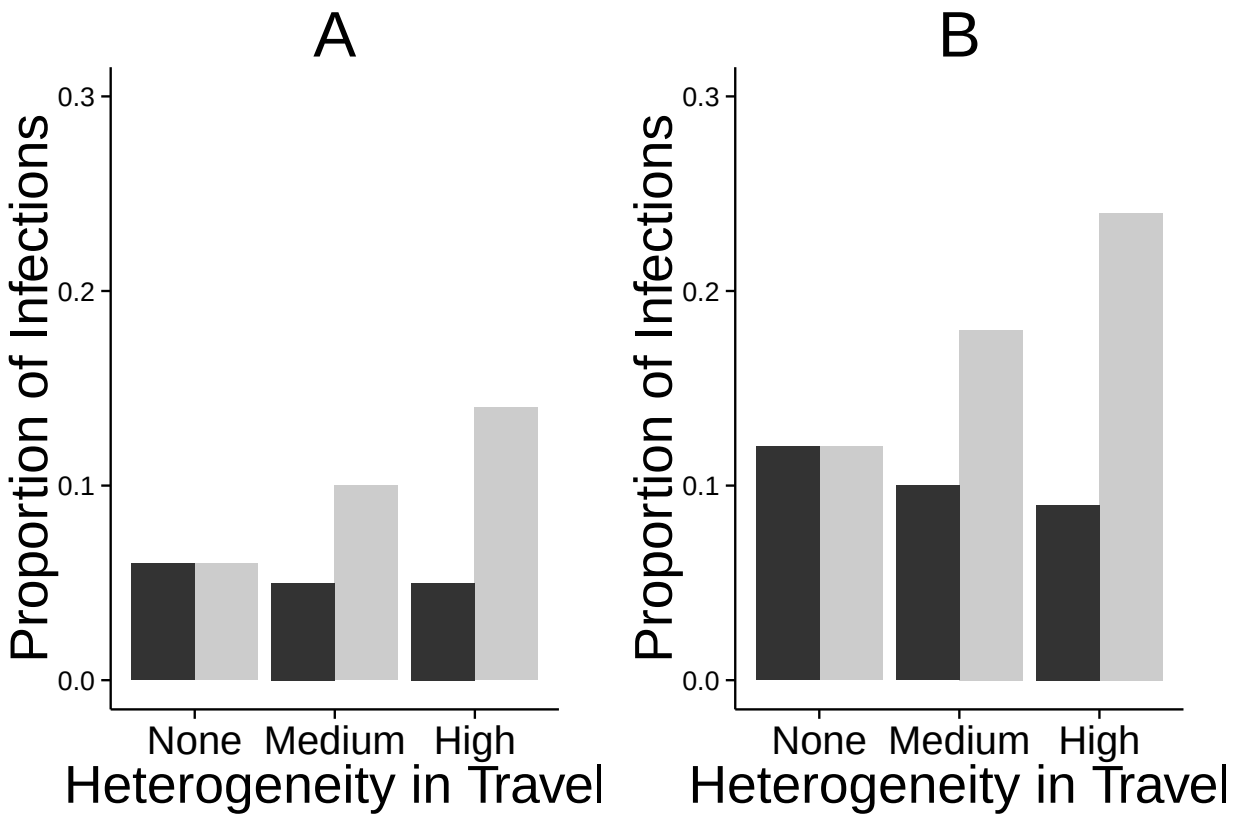

Figure 2.3: Attributable fraction for local transmission by heterogeneity in travel (adults vs. children) from the transmission model for A) low within-village transmission (Baseline $\mathcal{R}_0$=0.79) and B) high within-village transmission (Baseline $\mathcal{R}_0$=1.43). $\mathcal{R}_0$ increases slightly with increasing heterogeneity. Increasing heterogeneity was done using a proportionality constant where traveladult = ctravelchild and c takes values of 1 (none), 2 (medium), or 3 (high); c=1.73 in the study data. Adults are shown in black and children are shown in gray.

Table 2.5: Predicted cumulative incidence per 1,000 people by remoteness and study year. Models were parameterized to reflect data collected in our field site over time and from the literature regarding the three disease processes: travel to the city (by community, age group, and study year), transmission rate (by community and age group, but not time), and recovery rate (estimated from literature at five days). Community population size and the fraction of children under 5 years of age were also estimated for each study year using survey data and were incorporated into our model. The population of the city was fixed at 5,000 for all study years. All parameter values are shown in the supplement section A.5 along with relevant derivations.

Remoteness	2003	2007	2010	2013
Close Villages	13	24	13	28
Medium Villages	12	16	16	31
Far Villages	7	10	10	14

2.4 Discussion

Areas with changing transportation infrastructure undergo demographic change that can impact community structure and increase infection risk. In our analysis, this increased risk is largely driven by the city, highlighting the likely importance of population centers for transmission into remote regions. Furthermore, we show that increasing adult travel frequency can indirectly increase risk for those that travel less (e.g., children) and affect the fraction of infections that are acquired locally for both age groups.

2.4.1 Demographic change

Medium and close villages had higher rates of travel than far villages at baseline (2003). Over time, travel rates increased across the region. Similar increases in travel in our transmission model led to increased risk of infection. Although we parameterize our model to reflect rotavirus, our results naturally extend to other pathogens and demonstrate how road construction can lead to increased regional risk of enteric infection and highlight the need to consider the broader health impacts of constructing roads and other projects like railroads that can increase opportunities for human movement.

Several demographic variables were found to be important predictors of travel; the effect of these variables did not appreciably change over time. Travelers were more likely to come from households with higher education and to be salaried or domestic workers. Individuals with salaried positions were the most likely to engage in out-of-region travel. The increased travel seen among domestic workers may reflect fewer time constraints placed on women who work inside the home. The tendency for young children under the age of five to travel more than older children (aged 5-13) may be because the youngest children are too young to be left behind in the villages when the mother is traveling.

2.4.2 Travel and disease transmission

Our transmission model found that increased travel led to increased infection; most of this risk to the village appears to be driven by the city. Although the city had a higher transmission rate than the villages, our sensitivity analysis revealed that the city's population size was the primary reason that it dominated transmission. Adults experienced the greatest increases in infection risk due to their more frequent travel, but children were also indirectly affected by the travel of adults. This result highlights the role of travelers in dissemination of disease to remote communities and shows that population centers can be important for driving risk in remote regions.

The increased travel observed across all villages over time indicates potentially increased disease transmission for all types of communities–including those in remote regions–as a function of changing mobility. In general, more remote communities tend to benefit from less infection pressure due to their relative isolation [94]. In our study, remoteness remained an important predictor of travel and infection throughout the study, despite ongoing development. Therefore, at least in our study region, the erosion of a remoteness advantage occurs over a time scale longer than 10 years. Thus, population-level proxies (i.e., gravity models, radiation models, etc.) for travel may be reasonable predictors of

average travel behavior as development proceeds for capturing the average travel behavior.

Because of demographic differences in travel rates, local interventions may have a greater impact on children than adults but targeting interventions in cities is likely to be the most effective way to lower regional risk. Future studies comparing prevalence of infection and diarrheal disease both by remoteness and over time in this region may provide further insights into how changing travel patterns affect risk.

2.4.3 Strengths and Limitations

There are a few caveats to our study. First, our remoteness variable was defined at baseline. If community remoteness confounds the association between the considered covariates and travel, there may be some residual confounding by remoteness in later years that is not captured by the baseline measure. Second, because of the lack of preexisting theory on determinants of remoteness, we had to rely on the change in estimate criteria to evaluate confounding and mediation.

Additionally, our estimate of $\mathcal{R}_0$* required an estimate of age at first infection. We used age-specific prevalence of infection as a proxy for age at first infection. Because people can be re-infected with rotavirus, this assumption may have led us to overestimate the average age at first infection and thus underestimate both transmission rates and $\mathcal{R}_0$*. However, sensitivity analysis demonstrated that i) our results were similar regardless of the actual value of $\mathcal{R}_0$* and ii) relative infectivity (city vs. community and adults vs. children) was more important than absolute infectivity for determining the source of cases. Because this overestimation of β is non-differential by remoteness and thus relative infectivity is unaffected, our main results should be unbiased. Furthermore, people are rarely infected more than twice, and subsequent infections are less likely to be symptomatic and thus less likely to transmit [3, 95, 96]. As such, any bias in the absolute infectivity is likely to be minimal.

These longitudinal data allow us to assess change over time and show that the covariates considered have stable relationships with travel. Additionally, the analysis adequately accounts for the clustered nature of the data. The timing of this study relative to the completion of a road construction project is also a strength because the data provided a useful natural experiment with a unique opportunity to answer these questions in a population with new road development.

Overall, these findings provide useful data and a theoretical framework for future studies investigating the determinants of movement in other populations, particularly areas that are rapidly urbanizing. Given that rural-urban connections are increasingly more common, the relevance of this work continues to increase [97].

2.4.4 Conclusions

Current methods for incorporating travel into disease transmission models generally rely on proxies for travel (e.g., gravity and radiation models) that are based on distance and population size, without any consideration of demographic heterogeneity [51, 75, 76, 83, 84]. Our results suggest that while these models may accurately estimate average travel, accounting for travel heterogeneity in transmission models may both improve model predictions and identify opportunities to target public health interventions. Our results also support the use of cell phone data to characterize mobility both because risk of infection for children is largely dependent on the travel patterns of adults and because young children often travel with their parents.

CHAPTER III

Modeling Environmentally-Mediated Rotavirus Transmission: the Role of Temperature and Hydrologic Factors

3.1 Introduction

Diarrheal disease is the fifth leading cause of mortality worldwide and the third leading cause of death among children under five. The World Health Organization estimated that diarrheal disease accounted for 2.9% of the Disability Adjusted Life Years (DALY) burden worldwide in 2015, making it the sixth leading contributor to the burden of disease [98]. A recent multi-country study on diarrheal disease etiology showed that rotavirus was one of four pathogens that together accounted for the vast majority of severe diarrheal disease [4]. The recent introduction of rotavirus vaccines has reduced the prevalence of severe cases but is less likely to impact the overall transmission rate, especially given variable coverage of the rotavirus vaccine in low- and middle-income countries, as well as lower vaccine efficacy in the developing world [61, 62]. As such, rotavirus is likely to remain an important pathogen for the foreseeable future.

Like other waterborne pathogens, rotavirus can be spread either directly from person to person or through contact with contaminated water. However, because of its low infectious

dose and high shedding rate among infected individuals, environmental control is generally not considered effective for rotavirus [99] and indirect transmission through water is often not considered when analyzing rotavirus transmission dynamics [3, 38–41]. Instead, researchers have generally assumed that transmission through contaminated water is negligible in comparison with transmission occurring within homes and from person to person (with fomites as an intermediary).

In the tropics, meta-analyses have estimated that every 1°C increase in temperature is associated with a 4%–10% decrease in rotavirus-associated diarrheal disease incidence [43, 44]. Given that there is some evidence of publication bias for the temperature rotavirus effect and that other seasonal processes may affect risk, the true association may be closer to the lower end of this range [44]. In temperate regions, researchers have seen that rotavirus is more common during the cooler weather months [45]. A number of studies have examined fomite transmission, and although rotavirus can survive on fomites for long periods of time (on the order of weeks), there is a lack of evidence that temperature has a strong effect on rotavirus persistence on fomites [23–27]. In LMIC tropical settings, an alternative environmental pathway, such as water, may contribute to the temperature driven incidence patterns. In temperate climates, seasonality in direct transmission is more likely to be explained by large seasonal shifts in relative humidity, with drier conditions promoting more direct transmission [23–27].

Various studies have attempted to predict the mechanisms by which increased temperature might alter risk of diarrheal disease by changing pathogen survival and growth dynamics, pathogen virulence, and human behavioral factors [1]. In general, the burden of diarrheal disease is expected to increase as climate change progresses [100]; however, there is a great deal of heterogeneity by pathogen [44]. For bacterial pathogens, increased temperature is related to decreased persistence but can also lead to expression of genes associated with increased virulence [101, 102] and the net effect of increased temperature is thus

expected to increase risk. In contrast, higher temperatures are only expected to decrease environmental persistence of viruses, including rotavirus [1], and thus are expected to reduce viral infection risks [43]. Although this body of work has helped establish the associations between environmental factors and disease transmission risk, studies are needed to better elucidate the mechanisms by which environmental changes like increased temperature affect risk [103].

In this study, we use a meta-analysis of experimental data to derive a statistical model for the relationship between temperature and rotavirus persistence in water. We then apply the resulting temperature sensitivity function within a mechanistic model of rotavirus transmission in two communities linked by a flowing body of water, examining the potential for waterborne transmission to disseminate rotavirus infections between communities and amplify cumulative incidence within communities. The hydrologic conditions in these two communities were parameterized to reflect the conditions found in our study site in rural, coastal Ecuador [68], which encompasses three rivers with distinct average flow velocities: the Onzole (17 km/day), the Cayapas (26 km/day), and the Santiago (69 km/day) [69]. We also assess how changing temperature might impact risk at both the community and regional scales and explore how local hydrologic conditions and access to safe water and sanitation services could impact this relationship. The aim of this analysis is to estimate how temperature-related changes in waterborne transmission, mediated by the waterborne decay rate, might help explain correlations between rotavirus incidence and temperature or season, especially in tropical settings.

3.2 Methods

An overall diagram of our conceptual framework and the flow of analysis is shown in Figure 3.1 below (adapted from [1]). In brief, we conducted a meta-analysis of studies that provided data on rotavirus die-off rates in water and developed a statistical model relating

temperature to die-off rate. We then used this model to help parameterize an ordinary differential equation (ODE) transmission model to examine the role of temperature on rotavirus transmission patterns. All analyses were conducted in R (v. 3.1.2). We integrated the ODE model using the `deSolve` package.

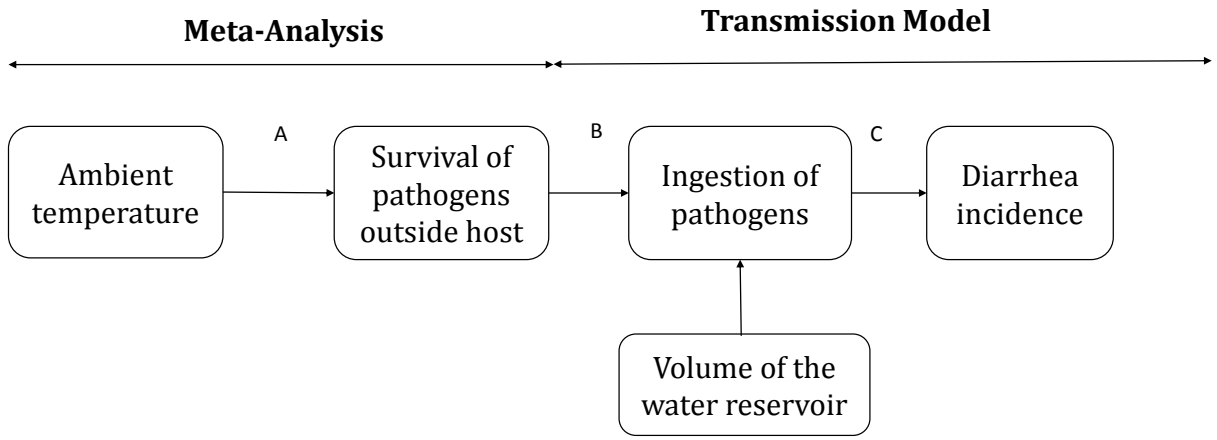

Figure 3.1: Conceptual framework of the analysis. The meta-analysis portion of our study informs arrow A, whereas the transmission model analysis informs arrows B and C. The transmission model also considered the role of the size of the water reservoir, which may vary based on the type of water source a community uses. This figure was adapted from [1].

3.2.1 Literature Review and Meta-Analysis

3.2.1.1 Data Collection

We retrieved a comprehensive list of studies that estimate rotavirus die-off in water samples at known temperatures by querying PubMed, Scopus, and Web of Science with the following search string:

(rotavirus OR 'rotavira') AND (temperature OR heat) AND ('die off' OR inactivation OR dieoff OR growth OR survival) AND (water OR aqueous OR aquatic OR marine)*.

Studies were included that: 1) reported results from at least one rotavirus experiment; 2) that took place in purified or treated water (i.e., reverse osmosis water, sterile distilled water, tap water, etc.), natural water (e.g., from lakes, rivers, creeks, etc.), post-treatment (i.e., tap water, filtered water, etc.), or simple aqueous solutions at neutral pH (i.e., phosphate-buffered saline (PBS)); 3) reported a steady water temperature that did not fluctuate more than 5°C during the course of the experiment; and 4) provided a clear explanation of methods used to assay rotavirus concentration over time. Studies conducted in more complex media (e.g., that added natural organic matter, humid sludges, etc. to aqueous media) were examined for control experiments conducted in pure water but were otherwise excluded. Also excluded were studies that reported only the time required for total inactivation of the rotavirus population.

Temperatures reported in die-off experiments were recorded directly from reviewed studies. Where mean rotavirus concentration was available in a figure or plot, data were extracted using DataThief software [104]. Additional information regarding rotavirus serotypes, water source, water treatment, initial viral concentration was also recorded.

3.2.1.2 Statistical Analysis

We determined the relationship between temperature and die-off rates in two steps. First, we estimated the die-off rate for each experiment. Second, we estimated the relationship between die-off rates in each study and the corresponding temperatures.

The rate of change in viral concentration is commonly modeled as a simple first-order decay process where $N(t)$ is the pathogen population at time t; N_0 is the initial population; and K is the first-order rate constant [105]:

$$N(t) = N_0 \times e^{-Kt} \tag{3.1}$$

We used this framework to model pathogen decay in our analysis. We also examined the data for evidence of tailing off, or biphasic decay [106], but few experiments appeared to exhibit biphasic decay and those that did lacked sufficient observations for us to estimate the parameters. We therefore assumed that the true pattern of decay was monophasic for all studies.

After obtaining an estimate of K in each experiment, we calculated the relationship between temperature and the decay rate of rotavirus in aqueous environments. We modeled the decay rate K of rotavirus as a non-linear function of temperature (T) and fit this model using generalized least squares, following prior published approaches used for other pathogens. In the equation below, K is the first-order rate constant at temperature T, K_{25} is the value of the rate constant at 25°C, and λ is a constant describing the change in decay rate for each 1°C change in temperature [107].

$$K(T) = K_{25} \times e^{-\lambda(T-25)} \tag{3.2}$$

3.2.2 Transmission Model

3.2.2.1 Model Structure

To explore the potential for hydrological transmission of rotavirus infections between sites, as well as the climate sensitivity of such an epidemiological system, we developed a susceptible, infectious, recovered (SIR) model representing two hydrologically connected communities at either end of a 10 km river reach. In our model, susceptible, infectious, and recovered individuals in community i are given by compartments S_i, I_i, and R_i, and transmission can be either direct—from contacts between susceptible and infectious persons,

which occur at rate β_H—or indirect—from ingestion of pathogens persisting in the aqueous environment (with contact rate β_W and per contact volume ingested ρ). The water compartments (W), described in more detail below, model the concentration of pathogens in the water. We assume that individuals within communities draw water from only one water source, such that the concentration in the water compartment is a reasonable proxy for exposure for all individuals within that community. We do not model human movement between communities.

We assume that infectious individuals shed pathogens into the environment at daily rate ϕ, but that the number of shed pathogens is attenuated prior to entering the water compartment by a combination of natural environmental decay and local sanitation and hygiene interventions (removing a fraction of pathogens c). In communities with built infrastructure, this parameter c reflects not only processes relating to runoff and decay in the environment, but also processes related to wastewater treatment, and thus most likely has a higher value. Larger cities may also have higher levels of attenuation due to their more complex topology. Upon deposition in the water, pathogens are diluted based on the total volume of the community's water source (V_i). Prior researchers have suggested that the per capita volume of water resources available may be smaller for large settlements with built infrastructure than for smaller communities, like those modeled here, where the size of the water reservoir is determined by natural rainfall and surface water flows, which often leads to greater dilution of shed pathogens [108]. We thus simulated across a range of available volumes by adjusting to produce water reservoir sizes that were consistent with a small pond (low dilution) to a medium-sized lake (high dilution) [109] (see supplement B.4.1 for additional details). The low dilution scenario also approximates the minimum the per capita rate of water availability targeted for larger cities with piped water [110]. Following our meta-analysis, we assume that the concentration of rotavirus in the water attenuates as a first order reaction with rate $\mu = K(T)$ (eq. ((3.2)). We model infections produced by ingestion of contaminated water using a linear dose-response function [111],

which we deemed adequate because the pathogen doses in our simulations fell within the linear range of the typical dose–response curves [112].

We modeled movement of rotavirus from upstream to downstream communities using a mixing cell hydrological transport model [113]. In brief, this model approximates the advection–diffusion equation for solute transport, modeling downstream flow through a series of compartments representing river reaches. While no diffusion of pathogens occurs in the upstream direction from point source inputs, the method produces a moving distribution of pathogen concentrations, with mean position at time after entering the system determined by the velocity of advection, and variance in pathogen position at time , i.e. rate of diffusion, proportional to the product of the advection velocity and cell length[113]. Thus, flowing systems and reasonably large spatial scales (i.e. where little diffusion in the upstream direction of pathogen sources is expected), our mixing cell approach closely approximates a solution to the advection diffusion equation. In our model, we used a total of 11 compartments (one for each village and nine intermediary), each representing a 1 km stretch of river, between which the rate of flow is determined by the flow velocity, which is given by ν. We assume that there is no sedimentation of pathogen between communities given the small virus particle size and the length of the river reach between communities. However, attenuation due to sedimentation could be approximated by assuming a slightly lower flow velocity. Pathogens entering the water at the upstream community are allowed to either move to the next compartment or decay, and no pathogens can be shed into the water between villages. More details about the mixing cell model, including comparisons with other methods, is in the supplement (section B.3).

The model diagram is shown in Figure 3.2, and the model equations are given below. Parameters used in simulations and their values are shown in Table 3.1.

$$
\begin{aligned}
\dot{S}_i &= -\frac{\beta_H}{N} S_i I_i - \frac{\beta_W \rho}{2} S_i W_i, \\
\dot{I}_i &= \frac{\beta_H}{N} S_i I_i + \frac{\beta_W \rho}{2} S_i W_i - \gamma I_i \\
\dot{R}_i &= \gamma_i R_i, \\
\dot{W}_1 &= \frac{(1-c)\phi}{V_1} I_i - (\mu + \nu) W_1, \\
\dot{W}_0^{\text{int}} &= W_1, \\
\dot{W}_j^{\text{int}} &= p\nu W_{j-1}^{\text{int}} - (\mu + \nu) W_j^{\text{int}}, \qquad 1 \leq j \leq 9, \\
\dot{W}_2 &= \frac{(1-c)\phi}{V_2} I_2 + \nu W_9^{\text{int}} - (\mu + \nu) W_2,
\end{aligned}
\tag{3.3}
$$

The transmission rate parameters β_W and β_H and the recovery parameter γ are assumed to be the same for both communities. In this simulation, the two communities have the same population size and available water volume. Flow velocities between the two sites are reasonable approximations of the hydrological characteristics of our study site in rural, coastal Ecuador [68], which encompasses three rivers with distinct average flow velocities: the Onzole (17 km/day), the Cayapas (26 km/day), and the Santiago (69 km/day) [69].

3.2.2.2 Model Analysis

All model simulations were seeded with one infectious case in the upstream community and were run for 365 days. Both communities were assumed to have a population of 1,000. To better understand the threshold and equilibrium features of our model, we calculated the basic reproduction number, $\mathcal{R}_0$, using the next generation method [119, 120]. Briefly, $\mathcal{R}_0$ captures the expected number of secondary cases produced by one case during its infectious period after being introduced into a completely susceptible population [120]. If this number is greater than 1 (e.g., greater than replacement) the outbreak will grow and if it is less than 1 the outbreak will die out. Because no pathogen is in the water at the

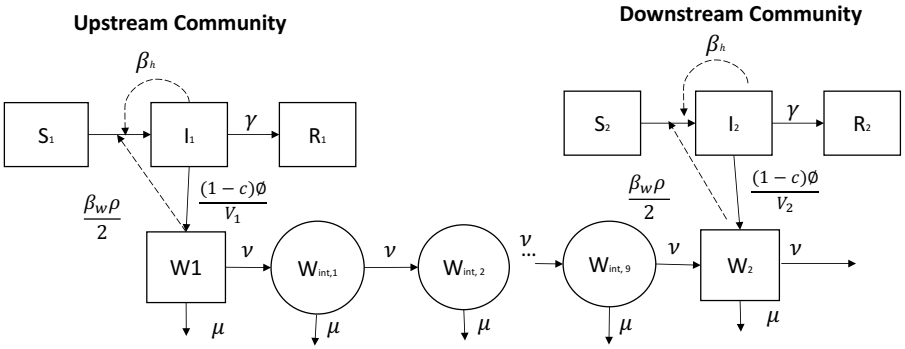

Figure 3.2: Rotavirus transmission model diagram. Dashed arrows represent transmission events and solid arrows represent transition of people or pathogens between model compartments. Villagers and city residents are classified as susceptible (S), infected (I) and Recovered (R). We also track contamination in the village water source (W). Susceptible people can become infected by either direct person to person transmission or indirect person-water-person transmission within their own community. Pathogens in the local water source can travel from the upstream to the downstream community through the water source. Pathogens in water sources can also decay in each community or between communities. The model parameters are defined in Table 1.

Table 3.1: Parameter estimates, units, and their sources. Water volume calculations and details regarding the scaling of flow velocity are shown in the supplement sections B.3.3 and B.4.1

Parameter	Symbol	Units	Value	Source
Transmission rate from human to human	β_H	1/day	[0, 0.6]	Varied
Transmission rate from water to human	β_W	water contact/day	3	[114]
Recovery rate	γ	1/day	0.2	[115]
Median infectious dose	n_{50}	ffu	10	[38]
Shedding rate	ϕ	n_{50}/person/day	2,605	[38, 116–118]
Attenuation of shedding	c	–	[0, 1]	Varied
Water consumed per day	ρ	volume/water contact	0.4	[114]
Population size	N	people	1000	Assumption
Water volume	V_i	L	$[10^5, 10^9]$	Varied
Pathogen decay rate	μ	1/day	[0,1]	$K(T)$
Flow velocity	ν	km/day	[0, 69]	[69]

disease free equilibrium, the $\mathcal{R}_0$ for the upstream community, the downstream community, and the system as a whole are all equivalent.

Amplification. To determine if water could amplify infection by changing the size of the outbreak, we estimated the overall attack ratio, that is, the cumulative incidence, as a function of the strength of each pathway and estimated how cumulative incidence changed as temperature changes impacted rotavirus decay rates. We also assessed how local hydrologic conditions and attenuation of shedding might impact risk (defined by cumulative incidence at the end of the simulation) both overall and from water sources alone.

Dissemination. We defined dissemination as 5% of the population in the downstream community being infected by the end of the simulation (365 days). We investigated how dissemination might depend on temperature, river flow velocity, and dilution for biologically plausible parameter values (see table 2 for credible parameter ranges). Because even established thresholds ($\mathcal{R}_0 = 1$) may not be stable in real world settings [119], i.e., outbreaks may be possible for $\mathcal{R}_0 < 1$ and knowing the parameters is alone is not always

sufficient to predict whether or not an outbreak will occur, we also analyzed the stability of this threshold for biologically plausible changes in initial conditions (initial outbreak sizes up to 10%).

We recognize that human travel is also an important disseminator for rotavirus and is needed to explain spread of disease to upstream communities that would not otherwise become infected through flowing waterways. However, in this analysis, we aimed to determine if water was also a sufficient disseminator to downstream communities. Thus, we acknowledge that waterborne dissemination is not necessary to spread disease between communities but aimed to determine if it was also a sufficient pathway.

For both dissemination and amplification analyses, we used the fitted model equation from the meta-analysis to calculate decay rates at different temperatures, simulating across the range of decay rates seen in the literature.

3.3 Results

3.3.1 Meta-Analysis: Temperature and Die Off

Our literature search returned 57 articles. Of these articles, we identified 9 that met our inclusion criteria containing a total of 39 experiments for use in our analysis. These studies are summarized in Table 3.2. More descriptive information is in the supplement section B.1. We used data from these 39 experiments to estimate the die-off rate (/day) for each study. The decay rates ranged from 0.008 to 0.996 per day, with a median decay rate of 0.056/day (corresponding to a mean survival time of 18 days).

Our pooled data across all 39 experiments support the hypothesis that the decay rate of rotavirus is temperature dependent, with higher temperature leading to increased die-off (λ=0.095, SE=0.018, $p < 0.01$). However, due to the exponential relationship between

Table 3.2: Studies included in the meta-analysis. The phrase 'combined' is used for simple aqueous solutions at a neutral pH, such as phosphate-buffered saline (PBS). The phrase 'purified' is used for sterile, distilled water. The phrase 'treated' is used for water taken from raw water or sewage that was treated to remove microbes prior to analysis.

Author, Year	No. Experiments	Temperature	Media Type	Ref
Espinosa et al., 2008	4	15, 25°C	Natural, Combined	[28]
Ward et al., 1986	9	4, 16, 23, 27, 29, 37°C	Natural	[29]
Raphael et al., 1985	12	4, 20°C	Natural, Treated	[30]
El-Sanousy et al., 2014	3	4, 22, 35°C	Natural	[31]
Hansen et al., 2007	1	22°C	Treated	[32]
Sattar et al., 1984	2	4, 20°C	Treated	[33]
Höglund et al., 2002	2	5, 20°C	Combined	[34]
Chung et al., 1993	2	5, 25°C	Natural	[35]
McDaniels et al., 1983	4	8, 26°C	Purified, Treated	[36]

temperature and die-off, the effect of a given change in temperature was small below 20°C (Figure 3.3). It is known that competition between microbes can influence decay rates and some of these experiments were conducted in natural water sources that could contain other microbes. To account for the possibility that the association between temperature and decay could be confounded by microbial competition, we also adjusted for this variable, with natural water sources being the only water type likely to contain other microbes. As expected, experiments conducted in natural water sources that could contain other microbes had higher decay rates (χ_{nat}=1.21, SE=0.327, $p < 0.01$). Although the effect estimate for temperature attenuated slightly after adjusting for this possible competition effect, the association remained significant (λ=0.078, SE=0.016, $p < 0.01$). We elected to use the adjusted results. We also ran the simulations using the unadjusted model as a sensitivity analysis, and the results were similar (see supplement section B.2 for unadjusted results).

The adjusted, fitted regression is:

$$K(T) = \exp\left(-2.51 + 1.21\chi_{\text{nat}} + 0.078(T - 25)\right) \tag{3.4}$$

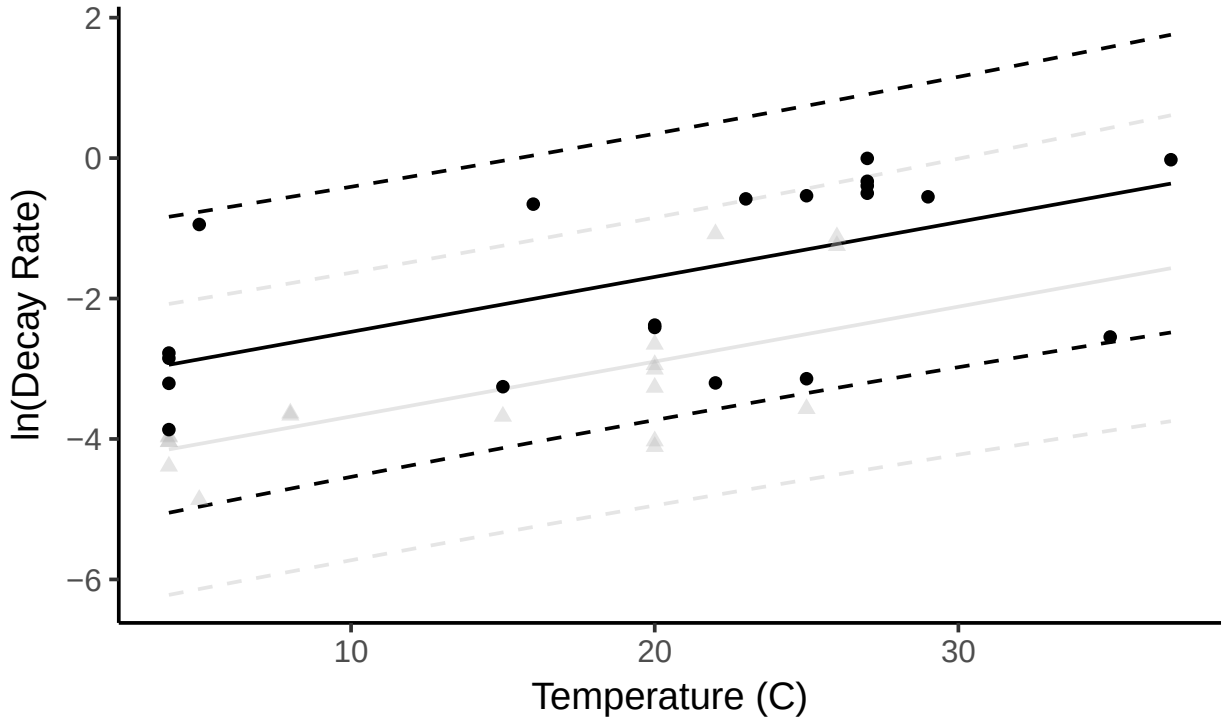

Figure 3.3: Temperature and Decay Rate with prediction intervals from the adjusted model. Data, predicted decay rates, and prediction intervals are shown in black for natural water sources and in gray for all other water sources. The predicted decay rates for each temperature are shown in solid lines and the 95% prediction intervals are shown in dashed lines for both types of water sources. For a graph of these decay rates on the linear scale, see supplement figure B.5.

There was no significant interaction between water type and temperature in our dataset.

3.3.2 Transmission Model

3.3.2.1 The Basic Reproduction Number

Using the next generation approach, we derived the following expression for $\mathcal{R}_0$ (see supplement):

$$R_0 = \frac{\beta_H}{\gamma} + \frac{0.5\beta_W N\rho\frac{(1-c)\phi}{V}}{\gamma(\mu+\nu)} \tag{3.5}$$

$$= R_{0,H} + R_{0,W} \tag{3.6}$$

The basic reproduction number can be written as a sum of two submodel reproduction numbers $\mathcal{R}_{0,H}$ (the $\mathcal{R}_0$ from human-to-human transmission) and $\mathcal{R}_{0,W}$ (the $\mathcal{R}_0$ for water-to-human transmission).

This expression leads to several insights. First, when all transmission occurs from person to person ($\beta_W = 0$), the reproduction number reduces to the SIR form of β/γ, which is the general formulation assumed by other rotavirus transmission models. Second, for water-human transmission, the river flow velocity impacts how strong of an effect temperature has on $\mathcal{R}_{0,W}$. For flowing water systems ($\nu > 0$) temperature related die off ($\mu = K(T)$) has a negligible effect whereas in standing water ($\nu = 0$), temperature related die off becomes more important for determining transmission potential. Third, the river flow velocity also impacts persistence in the local water supply. In flowing systems, the speed of flow is generally much faster than the speed of pathogen decay. As such, the length of pathogen persistence in the *local* water supply is much more strongly affected by hydrological flow dynamics. In standing water systems, the length of pathogen persistence in

the environment is entirely driven by temperature. Even in standing water, however, the most important contributor to water-human transmission ($\mathcal{R}_{0,W}$) is the size of the water reservoir (V) relative to the effective shedding rate $(1-c)\phi$. Overall, the size of the water reservoir relative to the human population is the critical driver of the strength of water-human transmission as it determines the degree of dilution that occurs when pathogens are shed into the water.

3.3.2.2 Amplification in Water Sources

In our simulations, for flowing systems, water can only amplify infection when the available volume of drinking water sources is low, because residence times in the local water supply are too short for pathogen concentrations to accumulate locally. However, for standing water, temperature can be an important determinant of community risk at moderate-to-high volumes, with higher temperatures reducing the strength of indirect transmission (Figure 3.4) and leading to lower cumulative incidence (Figure 3.5).

The strength of the temperature effect depends on both the strength of direct person-to-person transmission and the size of the water reservoir, with temperature having a stronger effect when direct transmission is weaker and the per capita water availability is higher. When per capita water availability is 1,000 L, temperature only affected risk at higher temperatures, with an average risk difference per degree Celsius of 1.0%, 0.3%, and 0% for temperatures between 20°C and 37°C for $\mathcal{R}_{0,H}$ values of 1, 2, and 3 respectively. When per capita water availability was 10,000 L, temperature had the largest effect on risk at lower temperatures with risk differences of 2.4%, 0.5%, and 0.2% per degree Celsius between 10°C and 20°C for $\mathcal{R}_{0,H}$ values of 1, 2, and 3 respectively. Between 20°C and 37°C, these same risk differences were smaller, with values of 1.9%, 0.3%, and 0.1% respectively.

For standing water at moderate dilution, the amount of sanitation coverage needed to reduce $\mathcal{R}_{0,W}$ below 1 varied by temperature, with higher coverage needed to interrupt

waterborne transmission at lower temperatures. When dilution was high, $\mathcal{R}_{0,W}$ was always less than 1 irrespective of the degree of attenuation of shedding. For flowing water, $\mathcal{R}_{0,W}$ was only greater than 1 when dilution was low, and the amount of sanitation coverage needed to reduce $\mathcal{R}_{0,W}$ below 1 depended on the flow velocity, with less coverage needed in faster flowing rivers (Figure 3.4).

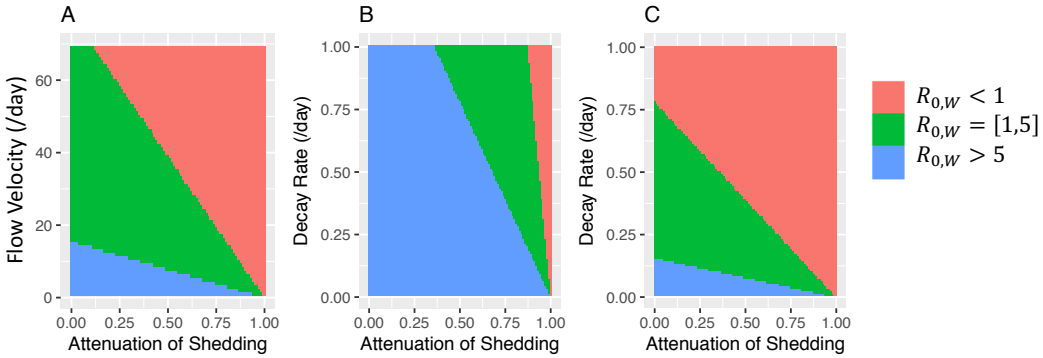

Figure 3.4: The strength of indirect, water-human, transmission ($\mathcal{R}_{0,W}$) for A) flowing water ($\nu > 0$) at low volume (10^5 L), B) standing water ($\nu = 0$) at moderate volume (1,000 L/person, 10^6 L total volume), and C) standing water ($\nu = 0$) at moderately high volume (10,000 L/person, 10^7 L total volume). For flowing water, the flow velocity (ν) is more important than decay rates, so the sensitivity analysis is shown as a function of flow rates (ν) and attenuation due to WASH interventions and natural processes (c). For standing water, the water is not flowing so decay rates (μ) and attenuation of shedding (c) become important for waterborne disease dynamics. For flowing water, the pathogen decay rate was set to 0.056, the median decay rate found in the studies included in our meta-analysis.

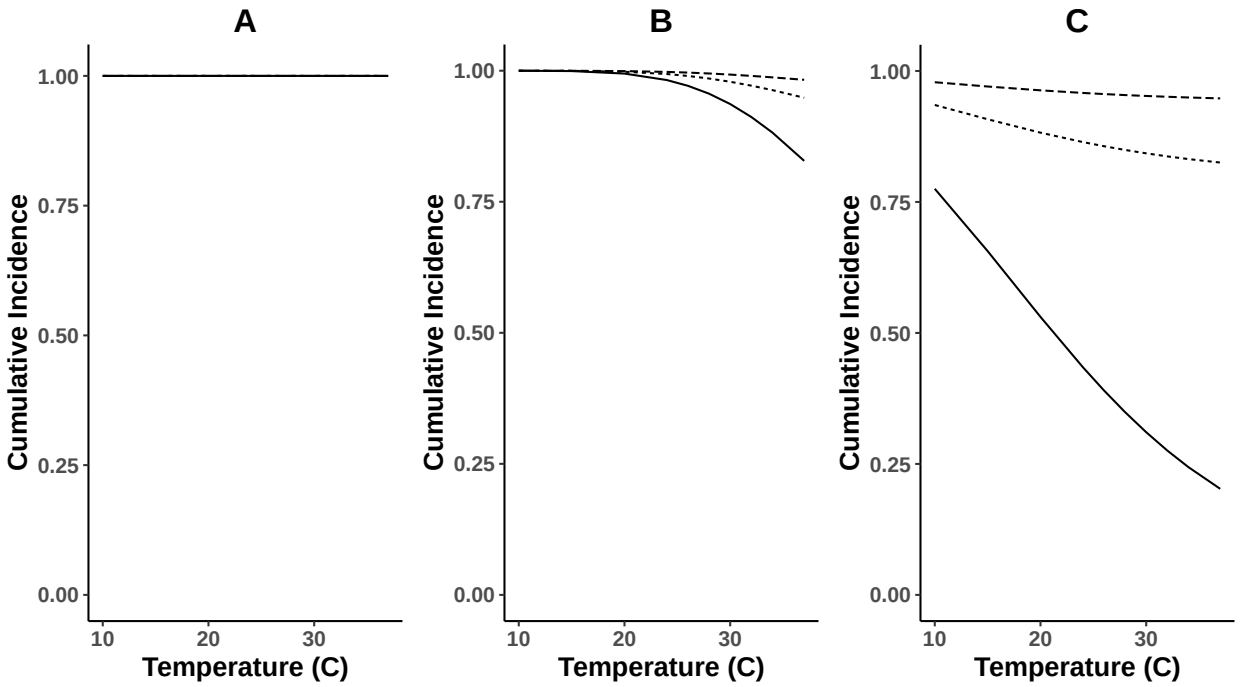

Figure 3.5: Cumulative incidence (fraction) for different temperatures and strength of direct transmission. $\mathcal{R}_{0,H}$ values of 1, 2, and 3 are shown in solid, dotted and dashed lines respectively for A) standing water ($\nu = 0$) at low volume (10^5 L), B) standing water ($\nu = 0$) at moderate volume (1,000 L/person, 10^6 L total volume), and C) standing water ($\nu = 0$) at moderately high volume (10,000 L/person, 10^7 L total volume). For all three panels, attenuation of shedding due to WASH interventions and natural processes (c) is set to 0.9. The maximum effect of temperature occurs at this level of attenuation and high dilution, with $\mathcal{R}_{0,W}$ decreasing from over 0.98 to 0.11 as temperature increases from 10 to 37° C for standing water. For flowing water, temperature has no effect on cumulative incidence for any volume because flow is much faster than decay and $R_{0,W}$ only changes from 0.05 to 0.04 as temperature increases from 10°C to 37°C.

3.3.2.3 Dissemination through water

In order for an outbreak to move downstream through water sources, we found that $\mathcal{R}_0 > 1$ at the location of outbreak seeding was a sufficient condition for downstream outbreaks, as long as the transmission term from water was positive ($\beta_W > 0$). However, this condition was not necessary when the water was highly concentrated (due to low volume and slow flow rates) and when the initial size of the outbreak was large.

When stream volumes are low, our model suggests that dissemination can occur even if the combined reproduction number (direct + indirect) is less than 1. This behavior is consistent across a wide range of die-off and flow velocity values, including both when there is no direct transmission ($\mathcal{R}_{0,H} = 0$, Figure 6) and when both pathways were activated ($\mathcal{R}_{0,H} > 0$ and $\mathcal{R}_{0,W} > 0$ but $\mathcal{R}_{0,H} + \mathcal{R}_{0,W} < 1$). In this last condition, the $\mathcal{R}_0 = 1$ threshold occurs at a higher flow rate whose value is proportional to the strength of direct transmission (Figure 6). At higher water source volumes above 100,000 L (100 L/person), there is no dissemination to the downstream community when $\mathcal{R}_0$ is less than 1. In summary, our simulations suggest that a sufficient condition for dissemination is that $\mathcal{R}_0 > 1$, and that there is a narrow parameter regime when $\mathcal{R}_0$ is slightly less than 1 in the upstream community where dissemination can also occur under low volume conditions.

When the system is driven by direct transmission, $\mathcal{R}_{0,H} < 1$ prevents dissemination, even when the initial size of theinfectious state in the upstream community isup to 10% of the population; i.e, $\mathcal{R}_{0,H}$is a stable threshold (Figure 3.7A). In contrast, when the system is driven by indirect transmission, the presence ofdownstreamoutbreaks and their size is sensitive to the initial sizeof the infectious state in the upstream communitywhen volume is low (Figure 3.7B); i.e., $\mathcal{R}_{0,W}$is not a stable threshold. The only time that $\mathcal{R}_{0,W}$is a threshold fordownstreamoutbreaks is when 1) the model is seeded with only one infectious casein the upstream community(sufficient for upstream outbreaks), and 2) volume is moderate to high (necessary for downstream outbreaks).

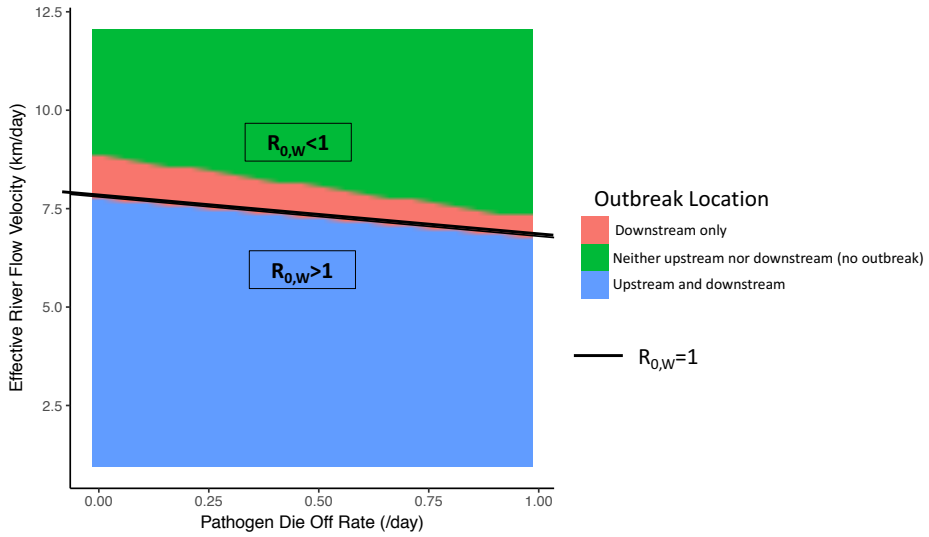

Figure 3.6: Location of outbreak (downstream and/or upstream community) as a function of pathogen decay rates and effective flow velocity. An outbreak is defined by a cumulative incidence of at least 5% after 365 days. The solid line corresponds to the situation when $\mathcal{R}_{0,W} = 1$. Above this line, $\mathcal{R}_{0,W} < 1$ and below this line $\mathcal{R}_{0,W} > 1$. For these simulations, the sub-model reproduction number for direct transmission was set to 0 and volume was set to 100,000 L.

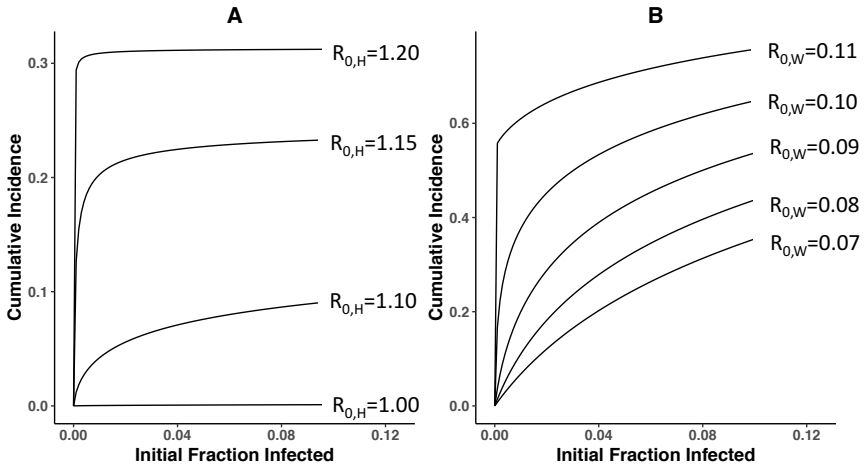

Figure 3.7: The effect of increasing the initial upstream outbreak size (fraction infected) on cumulative incidence in the downstream community when A) Most transmission occurs directly from human to human and the sub-model reproduction number for water is < 0.01. $\mathcal{R}_{0,H}$ is a stable threshold; downstream outbreak is driven by $\mathcal{R}_{0,H} > 1$; and B) All transmission occurs through the water and direct transmission is set to 0, under conditions of low volumes (100,000 L). Here, $\mathcal{R}_{0,W}$ is not a stable threshold; i.e., downstream outbreaks are possible for $\mathcal{R}_{0,W} < 1$.

Another condition that might prevent dissemination is if rotavirus die-off rates are higher than river flow rates. In our simulations, higher temperature resulting in higher die-off rates was not sufficient to prevent dissemination through waterways. That is because the die-off rates from our meta-analysis range from 0.008 to 0.996/day (corresponding to mean survival times from 1–125 days), which is lower than the river flow rates we observed in our Ecuador field site. Therefore, in practice, the parameter region where die-off rates are higher than flow rates probably only occurs for very slow flowing river systems where flow rates are less than 1 km/day. Furthermore, our simulations suggest that even when flow rates are slower than 1 km/day, dissemination can occur if volume is low.

3.4 Discussion

We find that environmental transmission of rotavirus through water sources may be an important source of risk in tropical climates at both the community scale (due to amplification) and regional scale (due to dissemination). The role of the environment in rotavirus transmission was affected by temperature, river flow velocity, WASH interventions, and the size of the water reservoir (which determines dilution). Increases in temperature were found to be associated with exponential increases in the decay rate of rotavirus in water sources. We would therefore expect temperature to only appreciably increase die-off in tropical locations where the water temperatures can easily exceed 20°C. However, even at lower temperatures small changes in ambient temperatures can have a larger impact on risk for communities drawing water from more dilute surfaces, such as lakes. Communities linked by flowing rivers could disseminate infection but their water supplies did not amplify within-community outbreaks unless dilution was low, whereas outbreaks in communities with standing water sources could be amplified by environmentally mediated transmission when dilution was low to moderate. In the real world, these patterns

may help to describe distinct disease dynamics among communities using rivers—as opposed to ponds and lakes—for water sources, as well as communities with riverine water sources with time-varying flow velocity and connectivity. Furthermore, even in communities served by flowing surface water sources, pathogens may accumulate in areas of slow flow along the river's edge [69]. Disease dynamics in communities where these zones serve as foci for interactions between the human population and the water source may exhibit qualities of both flowing and standing water communities as described by our model.

Prior researchers have attempted to understand how rotavirus may be influenced by temperature and other seasonal factors using both time series analyses and meta-analysis techniques. In general, these studies have found that rotavirus infections are more common in cooler, drier seasons with average effects of 4–10% decrease in incidence per 1°C increase in temperature [43, 44]. In our model, every 1°C increase in temperature leads to a total 0.1–2.4% decrease in incidence for standing water sources, which is at the lower end of the average associations between temperature and rotavirus seen in prior meta-analyses. The maximum temperature effect occurs when the reproduction number for direct transmission is close to 1. While this level of transmission is far lower than estimates of infection potential from prior models, in situations where the population is composed of both susceptible individuals and individuals with some immunity due to prior infection, the effective reproduction number is generally lower and may be close to 1 [121]. In situations like these, the effect of temperature on disease may be similar to our maximum temperature effect. There are a number of reasons that our results might not conform exactly to the results of these prior time series studies. First, because the relationship between temperature and decay is exponential, the effect of any temperature change depends on the baseline temperature. This means that larger changes in pathogen survival are expected in areas that are already hotter on average, such as the tropics. Second, many of the studies considered in these meta-analyses may also be subject to confounding by other unmeasured factors including differences in local hydrological conditions. Thus other seasonal processes are likely to

explain much of the remaining temperature association, and the average associations in the literature may be affected by publication bias [44].

The impact of temperature on disease transmission is likely to be different in temperate compared with tropical climates. In temperate regions, many factors in addition to temperature change seasonally that could cause a large shift in the strength of transmission, including seasonal changes in humidity. Thus, rotavirus seasonality in temperate climates is likely to be multifactorial and the result of many interdependent processes, which we do not explore here. While humidity has a large impact on the decay rate of rotavirus on fomites, temperature has a much smaller impact. The majority of the literature suggests that at high humidity conditions such as those found in the tropics, rotavirus survival on surfaces is likely to be long enough to sustain transmission for all temperatures [23–27]. Thus, while direct transmission may have a seasonal component, particularly in temperate climates, it is unlikely that the seasonality of rotavirus transmission in the tropics can be explained by temperature-related variations in fomite-mediated transmission alone [23–27]. Although we do not explicitly model other processes that might change seasonally in temperate climates, we do consider how population level impacts of changes in indirect transmission depend on the strength of direct transmission. Thus, our model may be thought of as modeling the variation in indirect transmission within a season in temperate climates.

Although our meta-analysis showed that higher temperatures were associated with faster decay, our model predicts that this effect of temperature would not be sufficient to prevent dissemination through water sources in sites similar to our study site in Ecuador. For communities with standing water sources, water also acted as an amplifier of local transmission. Even though the degree of coupling of infectious pathogens between the water supply and infectious people is probably lower in large cities that use standing water sources, the sensitivity of our model to the per capita water availability suggests that larger cities with

poor sanitation infrastructure where most residents get their drinking water from a single reservoir may be especially vulnerable to waterborne transmission. For communities using standing water sources, the effect of temperature was more pronounced, and higher temperatures led to smaller outbreaks due to the decreased persistence of rotavirus in the water supply.

3.4.1 Dissemination of rotavirus through flowing waterways

Together, our results suggest that water is likely to be a sufficient disseminator of infection in most flowing systems. Although it is not necessary for the overall reproduction number in the upstream community to be greater than 1 for dissemination to occur, in most contexts the sufficient condition ($\mathcal{R}_0 > 1$) is likely to be met either through direct transmission alone ($\mathcal{R}_{0,H} > 1$) or indirect transmission if there is increased contamination levels due to low water volumes ($\mathcal{R}_{0,W} > 1$). Because the overall reproduction number is a function of both direct and indirect transmission, it is sufficient for the sum of both submodel $\mathcal{R}_0$'s to be greater than 1. The submodel reproduction number $\mathcal{R}_{0,H}$ is likely to be greater than 1 in most contexts in the absence of focused interventions targeting this pathway, both because the shedding rate of rotavirus is extremely high relative to its median infectious dose and because rotavirus is persistent on a variety of environmental surfaces [23–27]. When $\mathcal{R}_{0,H} > 1$, the sufficient condition for dissemination will always be met, as long as water remains sufficient to disseminate enough pathogens to cause an outbreak in downstream communities. Even when direct transmission is not sufficiently high, low water volumes (increasing indirect transmission) or high initial outbreak sizes (increasing the probability of a breakthrough transmission event) may allow water to disseminate infection between communities.

3.4.2 Multiple pathways: direct and indirect transmission

While some seasonality patterns in rotavirus incidence may be partially explained by dynamic resonance related to the birth rate, many other factors may underlie this typical seasonal pattern of risk [39, 40, 122], including environmental variables like temperature. Because rotavirus is highly infectious, direct transmission alone may be sufficient to cause nearly universal infection of the susceptible population. However, in this analysis we have identified situations in which indirect transmission can affect the spread of disease through a watershed and may also be sufficient to cause large outbreaks. When two sufficient causes (i.e., environmental and direct transmission) are both met in a population[123], conventional epidemiologic studies that analyze case data alone may not detect this environmental transmission because multiple mechanisms are capable of explaining the observed dynamics [3, 4, 124, 125]. By explicitly considering this environmental route of transmission, we were able to better explore the potential importance of both pathways. Our model shows that while direct human-to-human transmission is likely to be more important for local amplification, indirect transmission may be sufficient to both sustain and spread outbreaks. Furthermore, if one assumes uninterrupted flow and surface water availability, waterborne transmission appears to be more important in cooler and drier seasons, when rotavirus persistence in the environment is high and the dilution effect is weak. In settings where certain assumptions of our model hold, i.e. that accumulation of and exposure to waterborne rotavirus can be adequately represented with a single water compartment and where access to safe water and sanitation is poor, inland cities with poor sanitation access are also expected to be especially vulnerable to waterborne transmission because of their relatively lower per capita water availability. In these settings, interventions like vaccination that impact both direct transmission and indirect pathways may be needed to reduce risk of infection within communities and to stop the spread of infection through watersheds.

Our model shows that the degree of intervention needed to interrupt water transmission depends on both local hydrologic conditions and ambient temperature. For standing water systems, less effort may need to be invested to prevent the accumulation of viruses in local water bodies when ambient temperature is high. For communities that utilize flowing surface water systems, targeting direct transmission alone may be sufficient to prevent disease spread as long as the size of the water reservoir is relatively large. When the size of the water reservoir is relatively small, the degree of intervention needed in flowing systems depends on the speed of water flow. Furthermore, if water is in fact an effective disseminator of rotavirus, restricting person-to-person contact may be sufficient to prevent the spread of rotavirus between communities for systems with greater per-capita water availability. In other types of communities, direct human-to-human transmission may be important for local amplification, but indirect transmission may sustain and spread outbreaks. This information can help public health practitioners identify optimal approaches to decrease environmentally mediated rotavirus risks in resource limited settings by considering local climatological and hydrological conditions.

We acknowledge that many questions remain about how rotavirus is transmitted: there is strong evidence both that the overall reproduction number $\mathcal{R}_0$ is high and that the environment can be mportant for transmission. In fact, the processes underlying rotavirus transmission are likely to be highly complex, with the potential for super-spreading and large heterogeneities in transmission by age [3]. In the United States, where drinking water and sanitation facilities are generally safe, rotavirus was still a major source of disease risk prior to the introduction of the vaccine, suggesting that water and sanitation interventions alone are unlikely to be adequate to interrupt transmission [126]. Because water remained an important source of risk in our model for communities resembling larger cities even when shedding was highly attenuated, our results are consistent with this line of research. Future work combining environmental sampling over multiple seasons with case data would be particularly useful to help validate our conclusions. Genetic

data could also help clarify the relative importance of different transmission routes. In addition, future studies on the effect of humidity on rotavirus persistence on fomites could help clarify other sources of seasonality in direct transmission.

3.4.3 Implications for trends in rotavirus risk under climate change

Given that global temperatures are projected to increase as climate change proceeds [127], many researchers have attempted to understand the relationship between temperature and risk of disease. In our model, we showed that one reason for the observed decrease in rotavirus infections at higher temperatures might be increased die off of rotavirus in water sources. Other pathogens, however, have different temperature signatures, and many studies have shown that risk of bacterial and protozoan diarrhea is, in fact, slightly elevated at higher temperatures [44]. However, these associations do not necessarily preclude increased die-off at higher temperatures for bacteria, but rather reflect the net balance of factors for bacteria compared with viral pathogens. For example, for many bacterial species, temperature activated virulence genes have been identified which may lead to increased risk of disease at higher temperatures even in the presence of faster die off [101, 102]. Some bacteria may also have higher replication rates at higher temperatures [44]. One implication, then, of our results is that the etiology of waterborne diarrhea cases might shift away from rotavirus and towards bacterial and protozoal infections as global mean temperatures continue to increase. On the other hand, rotavirus may be able to adapt to slowly increasing temperature, suggesting the need to study the temperature sensitivity of rotaviruses from different climates[128]. Another potential implication is that the waterborne transmission route for rotavirus may become less important as the average temperature increases, even if overall transmission does not decline. Given the non-linear relationship between temperature and die off, this shift is more likely in the tropics.

3.4.4 Conclusions

While some researchers acknowledge that waterborne transmission of rotavirus can occur, its importance is often downplayed relative to direct transmission [99]. Our model analysis reaffirms that direct transmission likely plays a central role in amplifying rotavirus infection within communities. However, we used rotavirus die-off data and prior meta-analyses to show that water is sufficient to both disseminate rotavirus infection between hydrologically connected communities, initiating local transmission cycles, and to amplify transmission under conditions of relative water scarcity in flowing systems and in standing water for a wide range of water availability. Our model indicates these environmental effects are most important in cooler seasons and, may be particularly important in larger cities with poor sanitation. These findings highlight the need to consider waterborne pathways in rotavirus control interventions, particularly in endemic regions in the tropics where the accumulation of virus in waterways may be more pronounced.

CHAPTER IV

Effect of Rotarix Vaccination in Rural, Coastal Ecuador

4.1 Introduction

Rotavirus is one of the top causes of severe diarrhea worldwide, particularly in children between 7 and 24 months of age [4, 10, 129]. In 2006, two vaccines were approved for use, Rotarix and Rotateq, with the objective of reducing severe disease [129, 130]. While both vaccines have shown similar effectiveness worldwide, Rotarix is the version used in Ecuador and in low resource settings in general [131]. However, this live, attenuated vaccine has to remain cold to be effective [132] and obtaining high levels of coverage is logistically difficult in remote regions [60–62]. To date, most studies on the efficacy and effectiveness of vaccination have come from randomized control trials and hospital based studies [55–58, 133–135]; data on how the vaccine performs in community settings is lacking [131]. Furthermore, no study has investigated Rotarix vaccine performance in Ecuador [131]. Here we examine Rotarix vaccine effectiveness in a community setting in rural Ecuador where there remain high barriers to vaccine coverage.

Beyond coverage barriers, developing country settings tend to have high genetic diversity of circulating strains [136], which results in repeated infection [18]. In part because of this greater diversity [59, 136], Rotarix vaccine efficacy has been shown to be inversely associated with overall childhood mortality, with weaker protection in countries with high

overall childhood mortality [131]. Because of the frequency of re-infection, rotavirus vaccines were expected to be most effective against severe rotavirus, with the goal of conferring protection similar to one natural infection [129]. A first natural infection generally provides strong homotypic protection against the infecting strain and subsequent infection provide broader protection against multiple strains [129]. On the other hand, rotavirus vaccines also provide some protection against severe all-cause diarrhea, potentially because rotavirus interacts synergistically with other diarrheal pathogens [37]. To our knowledge, the effectiveness of rotavirus vaccination against asymptomatic rotavirus infection and milder diarrhea disease has not yet been investigated.

In addition to showing strong efficacy among young children, data from vaccine trials has challenged prior understanding about how rotavirus is transmitted. In general, clinicians assume that infections occur throughout life but that symptomatic infections primarily occur among young children [16]. Because natural infection confers little protection against milder and asymptomatic infection, the vaccine was not expected to interrupt transmission a priori. However, a few studies have shown indirect effects on older populations that were too old to be vaccinated [65–67], suggesting that the vaccine may be lowering the overall transmission rate or potentially boosting antibody responses through circulation of vaccine strains shed in stool [137]. Most studies that have shown this effect have been conducted in the United States, where vaccine efficacy is highest and vaccine coverage is better than could be attained in many low-resource settings [16]. Whether or not such effects extend to low-resource settings is unclear. Determining if this indirect protection is present in low-resource settings is important because such effects would both increase the cost-effectiveness of vaccination and because they might stabilize the vaccine effect over time in remote regions whose transmission patterns are driven by periodic reintroductions of rotavirus by older, unvaccinated age groups [138].

Our objective in this analysis is twofold. First, we aim to determine if Rotarix implemen-

tation has been effective at reducing rotavirus infection and all-cause diarrhea in a remote region of Ecuador with high levels of endemic diarrheal transmission. Second, we hope to gain a greater understanding of how rotavirus is transmitted and explain counterintuitive finding from prior studies by quantifying 1) indirect effects of vaccination on older populations who were not vaccinated and 2) the impact of vaccination on asymptomatic rotavirus infection.

4.2 Methods

4.2.1 Data

This analysis built on data that were previously collected for a larger parent study on diarrheal disease in rural, coastal Ecuador [68]. This study collected yearly census information, active all-cause diarrhea surveillance, and 11 cycles of case control data (collected over a 10 year period). Some communities were dropped during the follow up period and others were added, but a total of 15 communities were followed for the full period.

We also collected all available vaccine records for children born in our study villages between 2008- 2013. We did not collect vaccine records prior to 2008 because the Rotarix vaccine had not yet been introduced at that time. Using date of birth, community of residence, and name, we linked these vaccine records to individuals who participated in our larger study. See supplement C.3 for more details about vaccination data collection.

We use this vaccine data along with ten years of case control data, 16 months of all-cause diarrheal disease surveillance, and the corresponding census data to assess the effect of the rotavirus vaccine on rotavirus infection, rotavirus-specific diarrhea and all-cause diarrhea in our study region. We collected these data between 2003 and 2013(Figure 4.1). The exact dates corresponding to the beginning of each study cycle are shown in the supplement C.2. All data collection protocols were approved by Institutional Review Boards at

University of Michigan and Universidad San Francisco de Quito. Details of the surveillance and case control outcome data as well as relevant covariates are described below.

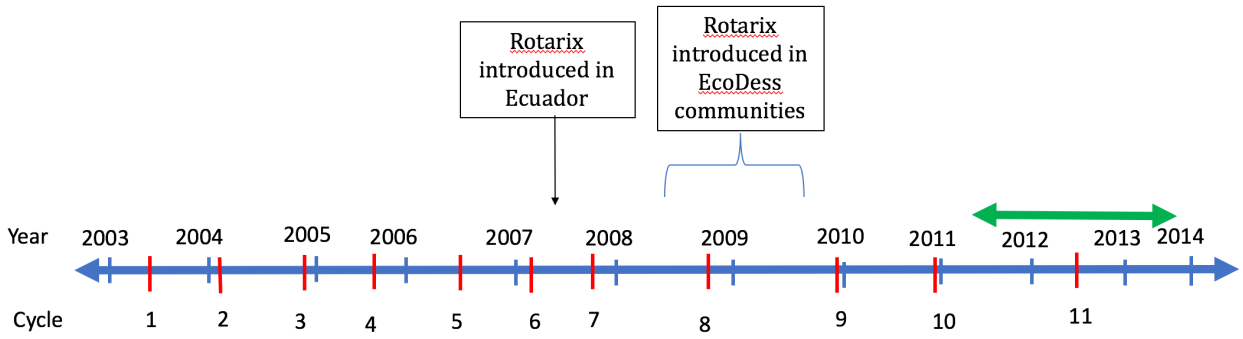

Figure 4.1: Data collection calendar, where each calendar year is shown in blue. Throughout the ten year study period, we conducted 11 serial, population-based, case control studies (timing of each is shown in red). The population based study design means that each individual in the community had a known probability of sampling. We conducted an active surveillance study between August 2011 and September 2013 (timing is shown in green), and a census in each cycle prior to the start of the case control study. Rotarix was introduced in Ecuador in 2007, but was not introduced in our study communities until late in 2008.

Surveillance outcome data. Between August 2011 and September 2013, we collected diarrheal surveillance data from all households every two weeks. However, there was a gap in data collection in July 2012 and from November 21, 2012–August 1, 2013. We recorded data for all cases of diarrhea.

Case control outcome data. Approximately once every 9 months from 2003 to 2013, field staff vis- ited each community and conducted a two week, prospective, population-based case-control study for symptomatic diarrhea. Due to the population-based design, we know the probability of sampling for each person in our study and can use this data to estimate population level prevalence and trends. In total, we have 11 cycles of case control data for each community. During the 14 day period when field staff were in the village, each case of diarrhea was identified prospectively. For the first seven study cycles, we collected one household control and two community controls for each case. For cycles

8-11, a random sample of 10% of the community population was sampled as controls at baseline. We obtained stool samples from each case and control and these samples were tested for rotavirus (and other pathogens).

Covariate data. Project staff used data collected at the start of the study to develop a metric for the remoteness of each village based on the cost and travel time required to reach Borbón, the largest town in the region [68]. This measure was subsequently categorized into 3 groups: 'close,' 'medium,' and 'far.'

Age was calculated using date of birth. For the surveillance analysis, we categorized age into $<$2 and $\geq$2 years based on age at entry into the cohort. Further disaggregation was not possible due to data sparsity. For the case control analysis, age was categorized into 3 groups: $<$1, 15, and $\geq$5 years.

We adjusted for household socioeconomic status using highest household education and household size. We chose to adjust for socioeconomic status using education both because it is generally considered to be an underlying cause of other socioeconomic indicators (such as occupation and income) and because other socioeconomic information was not available for the full population. Highest household education was the highest number of years of schooling reported by any individual of that household. Both variables were taken from census data. For the active surveillance, household characteristics were calculated from the census nearest to the child's date of birth. For case control analyses, socioeconomic indicators were taken from the census nearest to the case control cycle.

For the surveillance analysis, we also adjusted for BCG vaccination statues (yes/no) because it is generally administered at the same time as rotavirus and because it has been shown to have beneficial non-specific effects [139].

4.2.2 Analytic approach

To assess the effect of Rotarix vaccination in our study site, we conducted a three part analysis. First, we assessed the direct effect of the rotavirus vaccine on all-cause diarrhea illness using 16 months of active surveillance data (part 1). Second, we assessed the effect of community-level vaccine coverage (2 doses) on both rotavirus infection (including asymptomatic infection) and all-cause diarrhea, using ten years of case control data (part 2). Third, we assessed the effect of community level vaccination on the fraction of symptomatic diarrhea attributable to rotavirus infection using 10 years of case data (part 3). More details about the derivation of all vaccine effectiveness measures (defined using the convention of Halloran and Hudgens [140]) is in the supplement section C.1.

4.2.2.1 Part 1: Direct effects, cohort analysis

While both doses of Rotarix vaccine should be administered by four months of age, clinicians in our study region reported that the policy is to administer vaccination up to six months of age. We created a dynamic cohort of children beginning six months after birth, when both doses of the vaccine should have been administered. We then compared the rate of all-cause diarrheal disease illness by vaccination status (2 doses, 1 dose, 0 doses) using Poisson regression to estimate the direct effect of vaccination where 0 doses is the reference group. We calculated the direct effect of vaccination using the following formula:

$$VE = (1 - IRR) \times 100 \tag{4.1}$$

Based on our census records, there were 832 children born in our study communities between 2008 and September of 2013. Of these children, we have surveillance records for

773 children after 6 months of age. We have vaccine records for 361 of these 773 children and covariate information for 353 children. These 353 children (providing 5,945 person-weeks of data) were used in the multivariate regression model. Because vaccine records were not available for more than half of our study population, we also compared the demographic characteristics and diarrheal disease patterns of children with vaccine records and those without records to assess whether or not these groups were comparable. In general, children with records had higher rates of all-cause diarrhea than those without records. Because diarrhea is not seasonal in our study region, we did not adjust for seasonality in our model [44].

All statistical analyses were conducted in R (version 3.4). Weighted regression analyses used the package 'survey' and clustered logistic regression used the package 'gee.' Poisson regression analyses used the 'glm' function from the package 'stats'.

4.2.2.2 Part 2: Overall effects and indirect effects, case-control analysis

We compared the risk of rotavirus infection and all-cause diarrhea over time to assess whether the vaccine was associated with a change in transmission patterns in our study region. Seven communities in our study region had at least one positive stool sample for rotavirus in each cycle and we had vaccine records for six of these communities. While zero is our best estimate of rotavirus infection prevalence for communities who had zero positive stool samples for a given cycle, such estimates are subject to sampling error, particularly when prevalence is low, particularly because we only observing the population during a two-week time frame for each cycle. Additionally, in larger communities, this assumption could lead to bias in earlier cycles, when the control sampling strategy would have tended to capture a much smaller fraction of the population than in smaller communities. To avoid making this assumption, we restricted our longitudinal analysis to the six communities for which we had both vaccine data and positive stool sample data (n=2,489,

150 rotavirus positive). This approach has a predictable bias towards the null, because by design we could not have a 100% effectiveness estimate. In contrast, if we included all study communities we would expect our results to be biased but could not predict the direction a priori. To estimate confidence intervals for trends over time in the population and by age group, we used the Horvitz-Thompson variance calculation method.

We used logistic regression models, weighted by inverse probability of sampling, to compare the odds of rotavirus infection over time. We conducted this analysis separately by age group ($<$1 year, 1-5 years, and $\geq$5 years) and for all three age groups combined. For all models, community level coverage of two doses of vaccine was our exposure of interest (the proportion of children with vaccine records who received with two doses of rotavirus vaccine). We also considered a binary indicator (introduced/not introduced) but found that community level coverage was a better predictor and was a significant predictor of incidence patterns in the cycles after the vaccine was introduced. Thus, although average vaccination in the communities in our study overall provides some benefit (using a binary indicator), local coverage was an important determinant of community risk in later cycles. For this reason, we used community vaccine coverage as our exposure of interest in all regression models. We calculated overall vaccine effectiveness (VE) for all age groups separately and the population overall using the following formula [140]:

$$VE = (1 - OR) \times 100 \tag{4.2}$$

4.2.2.3 Part 3: Attributable fraction, case only analysis

In order to avoid changing the communities included in the analysis throughout the ten year period, which could bias estimates of population level trends, we restricted our analysis of symptomatic rotavirus to the 15 study communities that were followed for the study

period (n=830). We had vaccine coverage data for all fifteen of these communities. All fifteen communities had at least one case of diarrhea in every cycle. We used this data to assess whether or not the fraction of diarrheal cases attributable to rotavirus changed after the vaccine was introduced, using community vaccine coverage as our exposure of interest just as we did in part 2. Six of these fifteen communities were also included in the part 2 analysis. Thus, consistency between the symptomatic rotavirus regression in part 2 and the attributable fraction analysis in part 3 would increase our confidence that these findings generalize more broadly across the region. Because all symptomatic cases should be detected by our study design every case has a sampling probability of 1. Thus, these regressions were not weighted. However, for this symptomatic rotavirus regression analysis, we accounted for clustering by community using a generalized estimating equation (GEE) framework and an 'independence' correlation structure. We elected to use the GEE framework to facilitate population level interpretability [92].

4.3 Results

4.3.1 Descriptive statistics

4.3.1.1 Part 1: Cohort analysis

Differences between vaccinated and unvaccinated children are shown in Table 4.1. Children with vaccine records had higher rates of illness than children without vaccine records and a higher proportion of them became sick during the surveillance period. While not significant, there was a tendency for children with vaccine records to come from close/medium villages and for those without vaccine records to come from far villages ($p = 0.053$). Children without vaccine records also tended to come from houses with higher education ($p = 0.096$).

While some children may not have vaccine records because they were not vaccinated, some of these vaccine records may have been lost and many may have been vaccinated elsewhere, particularly Borbón, the nearest city in the region and location of the region's primary health center. Given that travel patterns in our study region are positively associated with socioeconomic status among adults [138], it is possible that the parents of children without vaccine records had higher income levels and brought their children to Borbón to receive vaccines. For this reason, we restricted our cohort analysis (part 1) to children with vaccine records.

Children who were vaccinated, in general, received both doses by six months of age (see table C.5 for details). Therefore, we assume that age upon entering the surveillance cohort (six months) is a good proxy for time since vaccination.

One community had very few records and the data quality was low, so this community was not included in the cohort analysis and is not shown in table 4.1. As a sensitivity analysis, we re-ran the regression analyses with this community added and while the confidence intervals were wider, the main results were unchanged.

4.3.1.2 Case control analyses (parts 2 and 3)

Vaccine coverage (exposure). For parts 2 and 3, we estimated vaccine coverage (our exposure of interest) among children with vaccine records only. We estimated that the average coverage was 74.6% for all cycles. Vaccine coverage was lowest in cycle 9 (66.4%) and highest in cycle 8 (78.8%) (see Table 4.2). At the community level, coverage estimates were variable, and ranged from 40.0% to 100%. The fraction of individuals with vaccine records was highest in moderately remote communities and among children with records vaccine coverage was highest in close villages. Overall, only about half of our study population had available vaccine records (see supplement section C.3 for vaccine record availability by community).

	No vaccine record	Vaccine record			
		Total	Zero Doses	One Dose	Two Doses
	(N=398)	(N=354)	(N=43)	(N=47)	(N=264)
Remoteness (%)					
Close	45.1 (180)	48.2 (170)	60.5 (26)	27.7 (13)	49.6 (131)
Medium	6.0 (24)	11.0(39)	11.6 (5)	8.5 (4)	11.4 (30)
Far	48.8(194)	40.8 (144)	27.9(12)	63.8 (30)	39.0 (103)
Age (yrs)					
Start of Surveillance	1.39 (0.85)	1.36 (0.80)	1.41 (0.70)	1.36 (0.82)	1.36 (0.81)
First Diarrheal Case	1.70 (0.84)	1.74 (0.86)	1.68 (0.44)	1.51 (0.73)	1.81 (0.95)
BCG Vaccination (%)	N/A	93.4 (330)	67.4 (29)	93.6 (44)	97.7 (258)
Male (%)	47.6 (191)	47.0 (166)	60.4 (26)	42.6 (20)	45.5 (120)
Household Size (n)	6.47(3.55)	6.63(3.62)	6.79(4.52)	6.72(3.38)	6.59(3.51)
Highest Household Education (yrs)	8.68 (3.61)	8.25(3.23)	8.56 (2.76)	8.22(3.52)	8.21 (3.25)
Overall Illness Rate(per 1,000 person-weeks)					
Diarrhea	21.4 (138)	27.5 (164)*	31.8 (26)	38.8 (33)	24.5 (105)
Non-diarrhea	3.6 (23)	6.7 (40)*	7.3 (6)	10.6 (9)	5.8 (25)
Any	25.0 (161)	34.3 (204)*	39.1 (32)	49.4 (42)	30.3 (130)
Person-time (weeks)	6436	5954	818	851	4285
Overall illness risk (%)					
Diarrhea	21.3 (85)	31.7 (112)*	34.9 (15)	42.6 (20)	29.2 (77)
Non-diarrhea	5.3 (21)	8.8 (31)	9.3 (4)	14.9 (7)	7.6 (20)
Any	24.1 (96)	36.3 (128)*	39.5 (17)	46.8 (22)	33.7 (89)
Length diarrhea episode (days)	3.02 (1.75)	3.22 (1.69)	3.08 (1.57)	2.97 (1.68)	3.33 (1.72)
Length first diarrhea episode (days)	3.36 (1.92)	3.36 (1.73)	3.29 (1.78)	3.40 (1.79)	3.40 (1.73)

Table 4.1: Surveillance Data Descriptive Statistics. Continuous variables are shown as mean(standard deviation), categorical variables and risks are shown as percent (n), and rates are shown as rate (number of events). * indicates a significant difference between children with and without vaccine records (comparing columns 1 and 2).

Remoteness	Cycle 8	Cycle 9	Cycle 10	Cycle 11
Close Communities				
Fraction with Records	40.2% (29/72)	56.6% (56/99)	52.5% (64/122)	32.9% (26/79)
Coverage of two doses	86.2% (25/29)	66.1% (37/56)	73.4% (47/64)	96.2% (25/26)
Medium Communities				
Fraction with Records	66.7% (6/9)	70.8% (17/24)	63.6% (14/22)	37.5% (6/16)
Coverage of two doses	50.0% (3/6)	64.7%(11/17)	85.7% (12/14)	100% (6/6)
Far Communities				
Fraction with Records	41.3% (31/75)	47.4% (46/97)	35.0 % (42/120)	53.6% (52/97)
Coverage of two doses	77.4% (24/31)	67.4% (31/46)	82.1% (32/39)	67.3% (35/52)
Overall Coverage	78.8% (52/66)	66.4% (79/119)	77.8% (91/117)	78.6% (66/84)

Table 4.2: Vaccine coverage over time by community remoteness

Time trends in all-cause diarrhea and rotavirus infection (outcomes). We estimated the prevalence of both all-cause diarrhea and rotavirus infection over time using the six communities included in the case control analysis. While older children and adults have the lowest per-capita risk of rotavirus infection (Figure 4.2A), they explain a substantial proportion of the infections due to their higher proportion in the population (Figure 4.2B). In this older age group, diarrhea was also a causative diarrheal pathogen (based on the odds ratio for diarrhea given rotavirus infection, see supplement section C.4). The total prevalence of rotavirus infection and all-cause diarrhea decreased after the vaccine was introduced, with young children $<$1 year of age showing strong reductions beginning in cycle 8 when the vaccine was first introduced and older age groups not showing this benefit until 1 cycle later (Figure 4.2A). Graphs of rotavirus infection (symptomatic and asymptomatic) and all-cause diarrhea over time are in the supplement section C.4.

Covariates for parts 2 and 3. In the case-control study, the prevalence of rotavirus infection in the general population and among cases of diarrhea declined after the vaccine was introduced (4.9% to 0.8% in the full population, 21.4% to 15.7% among cases, see Table 4.3). The prevalence of all-cause diarrhea also declined (2.2% vs. 1.6%). In the full population, the age distribution was similar throughout the study, but there was a tendency towards increased socioeconomic status as indicated by smaller household size and higher average

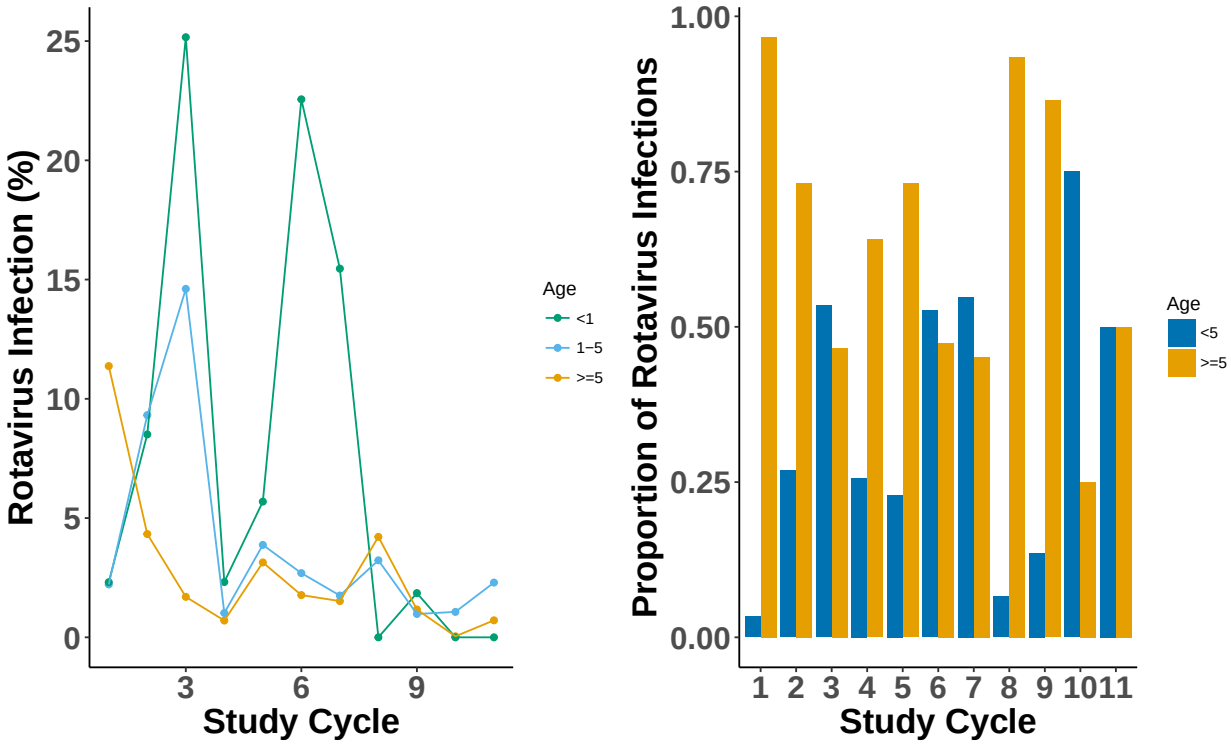

Figure 4.2: A) Percent of population infected with rotavirus by age group, B) Age distribution of rotavirus positive individuals (both symptomatic and asymptomatic) by study year. For panel A, children $<$1 year of age are shown in green, children between 1 and 5 years of age are shown in light blue, and older children ($\geq$5 years) and adults are shown in gold. For panel B, older children and adults are shown in gold and children under 5 years of age are shown in dark blue. The vaccine was first introduced in this region during cycle 8 (late in 2008).

	Case control analysis (part 2)		Case only analysis (part 3)	
	Pre-vaccine (N=20,660)	Post-vaccine (N=7,497)	Pre-vaccine (N=645)	Post-vaccine (N=185)
Symptomatic for diarrhea (%)	2.2(446)	1.6(122)	100 (case only)	100 (case only)
Male (%)	51.3 (10,617)	52.0 (3,902)	52.6 (327)	56.4 (101)
Age (%)				
<1	3.5 (713)	3.8 (286)	18.4 (114)	13.9 (25)
1-5	10.9 (2,260)	10.4 (777)	49.4 (306)	48.0(86)
>5	84.9 (17,548)	85.6 (6,432)	32.3 (200)	38.0 (68)
Household size (n)	6.75 (2.89)	6.34 (2.83)	6.37 (2.80)	6.70 (3.73)
Highest household education (yrs)	7.90 (3.38)	8.3 (3.35)	7.24 (3.32)	6.74 (3.33)
Remoteness (%)				
Close	66.1(13,653)	52.6 (3,944)	44.8 (289)	43.8 (81)
Medium	5.6 (1,158)	5.2 (390)	15.2 (98)	11.4 (21)
Far	28.3 (5850)	42.2 (3,164)	40.0 (258)	44.9 (83)
Positive for rotavirus (%)	4.9 (845)	0.8 (60)	21.4 (138)	15.7 (29)

Table 4.3: Characteristics of study sample before (cycles 1-8) and after (cycles 9-11) the vaccine was introduced. For part 2 (columns 2 and 3) all statistics are weighted by inverse probability of sampling and the total sample size (N) is the total population size of the six communities included in the analysis. For part 3 (columns 4 and 5), the statistics are unweighted and the total sample size (N) reflects the number of cases of symptomatic diarrhea for all 15 communities. For both analyses (columns 2-5), continuous variables are shown as mean(standard error) and categorical variables are shown as % (n).

household education in the later part of the study. Among cases of symptomatic diarrhea (part 3), the average age cases appeared to increase after vaccine introduction and individuals who did have symptomatic diarrhea appeared to have worse socioeconomic status as indicated by lower average levels of education and increased household size. Cases of diarrhea were also more likely to be male after the vaccine was introduced and to live in far villages.

4.3.2 Cohort analysis (part 1)

Among young children (under 2 years of age), rotavirus vaccination was associated with a decreased rate of all-cause diarrhea (Incidence rate ratio (IRR)=0.491, 95% CI: 0.289,

0.832). This IRR corresponds to a vaccine effectiveness of 50.9% (Figure 4.3 and Table S6). Among older children (over 2 years of age), the rate of diarrheal illness was similar regardless of vaccine status. While children who received one dose of rotavirus also tended to have lower rates of diarrhea than unvaccinated children during the first two years of life, this association was not significant (IRR=0.834, 95% CI: 0.453, 1.53), which most likely reflects the small sample size of this under-vaccinated group.

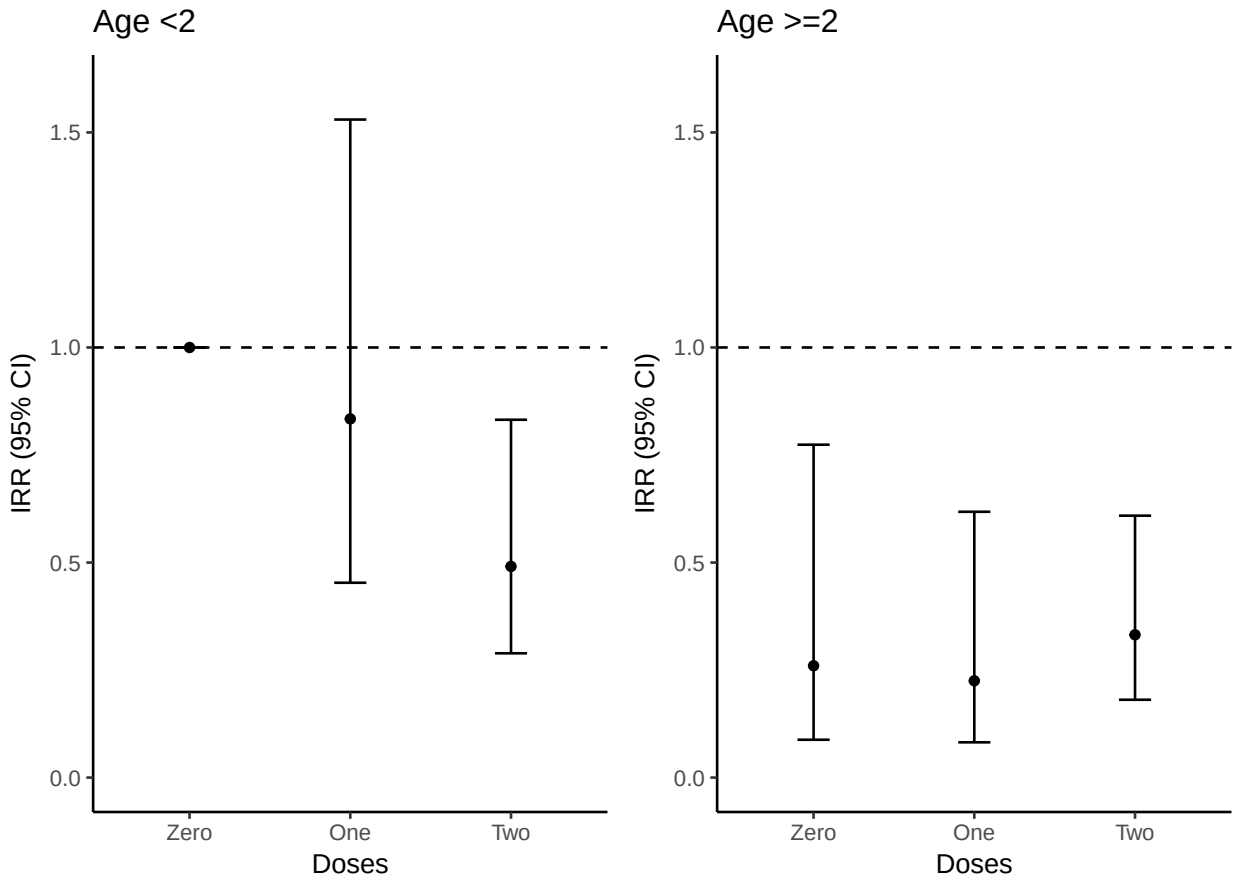

Figure 4.3: Effect of rotavirus vaccination on the rate all-cause diarrhea by age. Incidence rate ratios (IRR) are adjusted for community remoteness, gender, highest household education, household size, and BCG vaccination. For all rate ratios, the reference group is children less than 2 years of age who received zero doses of vaccine.

4.3.3 Case control analysis (part 2)

The community level vaccine coverage (among children who were vaccinated) was an important predictor of reduced rotavirus infection, with a 100% increase in coverage being associated with 0.118 times the odds of rotavirus infection in the fully adjusted model (Table 4.4), corresponding to an overall vaccine effectiveness of 88.2% (95% CI: 68.9%, 95.5%) (Table 4.5). While the strongest effect was observed among young children (<1 year) (Table 4.5), all age groups had reduced in rotavirus infection, including people who were too old to have been vaccinated (indicating the presence of indirect effects).

The overall effect of vaccination on asymptomatic infection was 92.8% (95% CI: 67.1-98.4%) (Table 4.5), which was stronger than the effect of vaccination on symptomatic rotavirus infection (VE=53.5%, 95% CI: 10.2-75.9%). For young children (<1 year), the effect estimate for symptomatic infection was 80.0% (95% CI: -17.8-96.8%). In later cycles, there were zero cases of asymptomatic rotavirus infection among children less than five years of age. Rotarix vaccination was associated with reduced all-cause diarrhea (VE=26.5%), but this association was only significant for young children (VE=73.3%, 95% CI: 3.3-92.5%). See table C.6 for a similar regression table for all-cause diarrhea.

4.3.3.1 Case-only analysis (part 3)

The community level vaccine coverage (among children who were vaccinated) was an important predictor of reduced rotavirus infection, with a 100% increase in coverage being associated with a 55% reduction in the odds of diarrheal cases being positive for rotavirus (see table 4.6). Despite the difference in communities included in the analysis, this result is similar to the reduction in symptomatic rotavirus in the full population based on the case control analysis (part 2), which found a 53.2% reduction. We also ran this model for each age group separately and found that the protective effect was strongest among the

	Unadjusted OR (95% CI)	Model 1 OR (95% CI)	Model 2 OR (95% CI)
Symptomatic for diarrhea	8.20 (4.94, 13.61)	8.54 (4.32, 16.8)	8.69 (4.31, 17.5)
Male	2.13 (0.993, 4.57)	2.19 (0.995, 4.82)	2.19 (0.991, 4.83)
Age			
<1	3.29 (1.04, 10.4)	2.71(0.727, 10.1)	2.74 (0.716, 10.5)
1–5	1.28 (0.531, 3.10)	0.766 (0.270, 2.17)	0.767 (0.270, 2.19)
≥ 5	Ref	Ref	Ref
Household size	1.13 (0.922, 1.41)	1.17 (0.955, 1.43)	1.15 (0.941, 1.41)
Highest household education	0.968 (0.842, 1.11)	0.928 (0.841, 1.02)	0.937 (0.848, 1.03)
Remoteness			
Close	Ref	Ref	Ref
Medium	0.942 (0.352, 2.52)	1.09 (0.402, 2.95)	1.09 (0.397, 2.97)
Far	0.202 (0.074, 0.551)	0.202 (0.074, 0.553)	0.241 (0.088, 0.662)
Community vaccine coverage	0.112 (0.042, 0.296)	–	0.118 (0.045, 0.311)

Table 4.4: Logistic regression of rotavirus infection (combined symptomatic and asymptomatic). All cells represent OR (95% CI). Reference group for categorical variables are noted above. OR for vaccine coverage compares 0 to 100%. The vaccine introduced variable is no for Cycles 1 ? 8 and yes for Cycles 9 –11. Model 1 is adjusted for symptomatic diarrhea, gender, age, remoteness, household size, and highest household education. Model 2 is adjusted for all variables in Model 1 plus community vaccine coverage.

Age group	Rotavirus Infection			All-Cause Diarrhea
	Any	Symptomatic	Asymptomatic	
<1	98.9% (63.6%, 100.0%)	82.9% (-4.0%, 97.2%)	100% (N/A)*	73.3% (4.4-92.5%)
1-5	82.9% (32.4%, 95.7%)	36.6% (-77.0, 77.0%)	100% (N/A)*	23.1% (-36.8-56.8%)
≥ 5	87.1% (57.8%, 96.0%)	53.8% (-16.4%, 81.7%)	89.4% (54.0%, 97.6%)	11.1% (-37.3-42.5%)
Total population	88.2%(68.9%, 95.5%)	53.2% (9.6%, 75.7%)	92.5% (66.1%, 98.3%)	26.5% (-5.7-48.9%)

*There were no asymptomatic rotavirus infections among children under five years of age after vaccine introduction.

Table 4.5: Overall effect of vaccination by age group, $VE = (1 - OR) \times 100$. Models for rotavirus infection (columns 2-4) are adjusted for symptomatic/asymptomatic status, remoteness, gender, household size, and highest household education. Models for all-cause diarrhea (column 5) are adjusted for all variables in columns 2-4 except rotavirus infection. Models for the total population (row 4) are adjusted for age.

	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Vaccine coverage (100% vs. 0%)	0.464 (0.293, 0.735)	0.453 (0.278, 0.739)
Male	0.855 (0.642, 1.14)	0.834 (0.625, 1.11)
Age		
< 1	1.68 (1.11, 2.53)	1.73 (1.06, 2.84)
1-5	0.642 (0.449, 0.917)	0.639 (0.448, 0.912)
≥ 5	Ref	Ref
Household size	1.02 (0.969, 1.08)	1.03 (0.964, 1.09)
Highest household education	0.992 (0.957, 1.03)	1.01 (0.975, 1.05)
Remoteness		
Close	Ref	Ref
Medium	1.34 (0.878, 2.05)	1.40 (0.943, 2.09)
Far	1.13 (0.852, 1.49)	1.18 (0.869, 1.59)

Table 4.6: Propensity for symptomatic diarrhea cases to be positive for rotavirus. Model adjusted for vaccine coverage, gender, age, remoteness, and household size.

youngest age group (<1 year), analogous to the results from the case-control analysis.

4.4 Discussion

In this study, we have shown that Rotarix vaccination has strong effectiveness on both rotavirus infection and all-cause diarrhea in a rural, nonclinical population in Ecuador despite modest vaccine coverage. These effects are strongest among young children who are generally at highest risk of severe disease. However, we also found indirect effects on both outcomes among older children and adults who could not be vaccinated, indicating indirect effects. These indirect effects are most likely due to suppression of the overall transmission rate, including asymptomatic infection, among all age groups.

Specifically, children < 1 year old were protected from illness (VE = 83% for symptomatic rotavirus infection and 73% for all-cause diarrhea; Table 4) at a higher level than previously reported in randomized clinical trials (RCTs) in Latin America. These RCTs found that rotavirus vaccination was associated with a 39%-42% risk reduction of hospitaliza-

tion for all-cause diarrhea [57, 141]. Our higher VE estimates for all-cause diarrhea can at least partially be explained by the fact that we were observing community-level illness of any severity. The overall effects for symptomatic rotavirus infection in our field site for $<$1 year olds (VE = 83%; Table 4.5) was in close agreement with similar vaccine trials in Latin America among young children (VE 81%; [57]).

While young children had the strongest protection against illness, all age groups were significantly protected from asymptomatic infection, including the oldest, unvaccinated age group (VE=92%). This result implies that Rotarix vaccination suppresses not only severe disease but also interrupts overall transmission and helps explain why we and others have observed indirect effects of vaccination in older age groups. We find that this benefit is attainable in low resource settings, despite modest levels of vaccine coverage (76% on average). The effect on older children and adults that we have found here may help reduce community exposure to rotavirus and ultimately improve health outcomes in younger children, who are at higher risk of severe disease. Such a reduction in infection might also help sustain the beneficial effects of rotavirus vaccination over time. Adults may play an important role as reservoirs for infection given their higher prevalence in the population, and may also be important for introducing pathogens into remote communities like these, which cannot sustain rotavirus transmission in isolation [138].

The suppression of asymptomatic infection (92%) also suggests that prior studies, which focused on clinical rotavirus infection, may have underestimated the total impact of vaccination. Evidence of a strong VE on infection is important for two reasons. First, there is increased evidence that repeated subclinical infections is linked to long-term sequelae of stunting and environmental enteric dysfunction [142]. Second, older children and adults can serve as reservoirs of rotavirus community-level transmission. Indeed, we found strong indirect effects on rotavirus infection in older children and adults.

The lag between vaccination of young children and protection of adults (Figure 4.2A)

suggests that the benefit to adults may be partially attributable to reduced amplification of rotavirus infection within communities by young children who were vaccinated for rotavirus [16, 138]. However, this delay may also reflect the time needed for pathogen decay in environmental sources. While we could not formally assess the fraction of transmission due to environmental transmission without environmental sampling data, we did find that rotavirus cases did not cluster very strongly within households (see supplement C.4.3), suggesting that most rotavirus infections do not occur in the household. Follow-up studies could combine environmental sampling with surveillance data to better assess the fraction of transmission due to this pathway.

Our vaccine data was collected retrospectively from secondary sources, which led to high levels of missing data for the all-cause diarrhea cohort analysis and forced us to use a smaller sample size. Because the children for whom we have vaccine records were sicker than children without vaccine records, our results may not be generalizable to children without vaccine records. However, given that the population level effects of vaccination were strongly dependent on our vaccine coverage estimates rather than vaccine introduction alone, it is likely that the actual vaccine coverage within our study communities is at least proportional to the coverage estimates from children with vaccine records. As a sensitivity analysis, we also estimated the effectiveness of two doses of rotavirus vaccine on all cause diarrhea in children under two years of age in the serial case control study (using community vaccine coverage as the main exposure), and found very similar reductions in the odds of all cause diarrhea (45%) (see supplement C.7). As such, it is likely that our estimates of vaccine coverage from the health post data are valid approximations of vaccine status in our study communities overall.

Although we did not conduct a formal analysis to explore to what extent the effect of vaccination was driven by interaction with other pathogens, the 73.3% effect estimate for younger children (Table 4.5) is larger than the fraction of cases that were attributable to

rotavirus for this age group for any cycle in the case control study. Thus, it is likely that the vaccine also affects risk of infection with other diarrheal pathogens, as others have also suggested [37]. We plan to conduct follow-up analyses to better address how synergistic effects might vary by co-infecting pathogens and whether or not this effect has led to a decline in stunting in our study region.

Because this is an observational study, we cannot guarantee that all of our exposure groups are comparable even after adjusting for relevant confounders. We plan to conduct follow-up analyses using simulation models to model counterfactual contrasts between vaccine groups explicitly, such that our exposure groups will be comparable by design. Finally, we are in the process of genotyping all rotavirus positive stool samples after the vaccine was introduced. We plan to use this data to assess whether or not the circulating strains of rotavirus have changed over time in response to vaccine introduction.

Rotarix vaccination can substantially improve health outcomes in remote communities with significant barriers to vaccination. Much of this effect is driven by suppression of overall rotavirus transmission at a population level–including among older age groups who were too old to be vaccinated. We find that the impact of rotavirus vaccination may be higher than previously estimated given its effect on milder diarrheal disease and asymptomatic infections. Because vaccination also reduces diarrheal disease morbidity, long-term health benefits, such as reduced prevalence of stunting, may also occur after vaccine introduction.

CHAPTER V

Conclusion

5.1 Summary

This dissertation expands our knowledge of the epidemiology of rotavirus transmission in low resource settings. In the introduction, I elaborate on the gaps in our understanding of rotavirus epidemiology. In the conclusion, I elaborate on how this dissertation addresses these gaps and how each one has advanced our understanding of rotavirus epidemiology. Specifically, in this thesis, we have examined the role of travel, the potential for environmental rotavirus transmission, and how the introduction of the rotavirus vaccine has affected transmission dynamics in rural Ecuador, impacting both direct and indirect transmission pathways.

5.2 Travel and pathogen introduction

5.2.1 Travel and urbanization: modeling the mean

With the rise of big data, many infectious disease researchers have begun using data from cell phone companies to estimate travel rates for disease transmission models for a variety of pathogens. However, such data may be difficult to obtain quickly during public health

emergencies [143]. Furthermore, these data are not always of high quality in remote regions, where cell phone towers are sparse. In such situations, simple approximations using gravity models and other similar approaches [51–53] are often used as a first approximation. Given the urbanization that is occurring in many countries worldwide, such approaches could be biased if infrastructure development does not occur in proportion to a region's remoteness. For example, if remote regions are particularly slow to acquire transportation infrastructure, communities might be more different than such approximations would predict. However, we find that in our study site in northern coastal Ecuador, the gradient in travel by remoteness has been maintained over time in a setting with ongoing road construction. Our results suggest that at least in our study region this concern of bias is unwarranted and that population level approximations should accurately capture average travel rates even in a setting with rapid change.

5.2.2 Adults and rotavirus transmission: the role of heterogeneity

At the same time, accounting for heterogeneity in travel patterns within communities may provide useful insights about sources and sinks of transmission. In our study region, we were able to show that age was an important determinant of travel patterns. Adults are key disseminators of infection throughout our study region and outbreaks cannot be maintained in isolation without the reintroduction of pathogens through human movement. In contrast, children are important for amplifying infection within communities and tend to be infected locally. These results are likely to be similar in other remote regions, and reinforce prior research suggesting that intervening in large cities is crucial for reducing regional risk of disease.

These two key results help contextualize the results from our vaccine analysis and explain why population level impacts of vaccination in remote regions might be greater than expected from vaccine coverage estimates alone. In our study region, we see substantial

vaccine benefits at around 75% vaccine coverage. Because the infection patterns of remote regions are strongly dependent on pathogen introduction, lower local levels of vaccine coverage may still provide a substantial benefit. Furthermore, because rotavirus infection in older children and adults is also indirectly reduced after vaccine introduction, the rate of rotavirus introduction through travel may decrease over time. This result enhances the population level impact of vaccinating young children and also helps to stabilize this benefit over time by reducing community susceptibility to small, sporadic introductions.

While our focus was on understanding rotavirus dynamics, subgroups may be important for spreading disease for other infectious diseases as well. For example, in Cambodia, forest workers are particularly likely to travel and may be important for spreading malaria between forest regions and larger cities [144]. In situations like these, both identifying and accounting for heterogeneity in travel can reveal opportunities to target interventions and help both predict and explain the impact of intervening on one group in isolation.

5.3 Water can both disseminate and amplify rotavirus infection

Our analysis in chapter three highlights that water can be an important transmission pathway for rotavirus, a pathogen that is commonly assumed to be transmitted directly. Waterborne transmission of rotavirus can partially explain temperature-dependent seasonality in rotavirus that has been observed in the literature, and also suggests mechanistic reasons why the association between temperature and rotavirus incidence varies by study. For example, the effect of temperature on decay also depends on the size of the water reservoir, the speed of water flow, and the baseline temperature. Because rotavirus is highly infectious and shedding is much higher than the infectious dose, both direct and indirect transmission is likely to be sufficient to cause substantial outbreaks. For this reason, traditional epidemiology studies that focus on surveillance data might overlook the environmental transmission pathway. Our work highlights the need for future studies to combine envi-

ronmental sampling with disease surveillance to assess how much transmission is occurring through both direct person-to-person and indirect person-to-environment-to-person pathways.

Given that adults are historically more likely to be infected from environmental sources than children and can function as a reservoir of infection in endemic settings [42, 145], interventions like vaccination, that impact both direct and indirect transmission pathways, are needed to reduce rotavirus disease at a population level. In addition to delays associated with vaccine rollout, environmental transmission for adults might contribute to delayed time between vaccine introduction and full benefit. The low level of clustering of rotavirus infections within households supports, but does not prove, this interpretation. While this analysis is a useful first step to establish the feasibility of water-mediated rotavirus transmission, more work remains to be done. Future studies that couple environmental sampling with epidemiological surveillance data would help reveal the importance of this indirect pathway and could also assess the relevance of within-household transmission. Additionally, although we found that temperature-related changes in rotavirus decay in water sources explains some of the seasonality of rotavirus in the tropics, our model does not capture the full association. Additionally, the water-borne pathway is less likely to explain seasonality in temperate climates, where water quality is far better. Future studies are needed to better understand the seasonality of rotavirus infection both in lower to middle income countries in the tropics and in high income countries, where seasonality is likely to have different environmental and social drivers.

5.4 Rotarix vaccination impacts rotavirus infection and all-cause diarrhea

We have shown that Rotarix vaccination is associated with substantial reductions in both rotavirus infection and all-cause diarrhea in a region of Ecuador with high barriers to vaccination. This benefit extended to asymptomatic infection and also impacted older age groups, who could not have been vaccinated. Because most prior studies have focused on clinical populations [55–58, 133–135], effects on milder rotavirus infection have not been well-established. Furthermore, we are not aware of any previous study has shown an impact of vaccination on asymptomatic infection. For this reason, the total impact of vaccination on overall rotavirus transmission may have been underestimated previously.

The suppression of rotavirus transmission has key implications for the long-term benefits of rotavirus vaccination. In general, the goal for rotavirus vaccination is to reduce the incidence of severe disease outcomes. However, if asymptomatic transmission is also suppressed, as our results suggest, Rotarix vaccination also ultimately reduces population level exposure to rotavirus, which can have huge morbidity benefits. Repeated infections with diarrheal disease pathogens, particularly in childhood, can cause long-term alterations in gut function [146]. This dysfunction can also lead to growth faltering [146]. Reduction in overall rotavirus circulation also helps explain why rotavirus vaccination has been associated with substantial benefits in unvaccinated populations [16, 65–67]. In our study region as well as in other endemic settings, rotavirus is not only a childhood disease but also affects adults [145]; these adults are also at risk of symptomatic diarrhea, suggesting that infection is not purely silent for this group. In addition to these direct benefits, adults may function as a reservoir for rotavirus and suppressing transmission from this group is important for reducing population risk.

Although these findings provide useful information about the effectiveness of rotavirus in

community settings many more questions remain. In addition to its impact on diarrhea, the suppression of rotavirus infection suggests that the vaccine may have broader impacts on other health outcomes like stunting that are related to diarrheal morbidity. Future research is needed to assess the potential long-term benefits of rotavirus vaccination. While we did not do formal synergy analysis for this dissertation, the vaccine effectiveness estimates we found are higher than would be expected if Rotarix only reduced risk of rotavirus. This result is consistent with other work by Bhavnani et al., which shows that rotavirus synergistically interacts with other co-infecting pathogens [37]. In our case control study, an average of 27.3% of cases of symptomatic diarrhea among children under two were positive for rotavirus before the vaccine was introduced. Therefore, even with a perfectly effective vaccine, only about 27% of all-cause diarrhea cases would be expected to be averted by the vaccine, which is far less than the 50.9% effectiveness we found among young children under 2 years of age. We plan to conduct follow-up analyses to better understand how the vaccine has impacted infection with other diarrheal pathogens in our study region.

While the rotavirus vaccine has shown high effectiveness initially, more work is needed to characterize how the vaccine might change the strains of rotavirus in circulation [16]. Several studies have shown that after vaccine introduction, the prevalence of the G2 serotype has increased, which one strain against which the rotavirus vaccine is less protective [147, 148]. In our study region, we previously showed that the circulating strains of rotavirus can change rapidly [149], which can lead to larger outbreaks. Furthermore, some of the strains circulating in our region could not be genotyped, suggesting that they may be new, emerging strains [150]. Although it is outside the scope of this dissertation, we are currently genotyping stool samples from Ecuador during the later part of the study after the Rotarix vaccine was introduced to determine how the circulating strains are changing.

5.5 Conclusions

The past few decades have been accompanied by dramatic progress in reducing rates of diarrheal disease. The global burden of disease project estimates that the burden of diarrheal disease overall dropped by 51.2% between 1990 and 2010 and the burden of rotavirus dropped by 55.8% [6]. However, diarrhea remains an important part of the disease burden worldwide. This dissertation has advanced our understanding of how rotavirus is transmitted. Hopefully this improved understanding can inform ongoing control efforts.

Our finding that adults are important for rotavirus transmission in endemic settings like our study context highlights the need to consider their contribution to overall dynamics. In particular, adults are important for disseminating infection between remote communities due to their higher rates of travel [138]. Given that adults are important for rotavirus circulation in endemic settings like our study context, it is encouraging that we find that the Rotarix vaccine also reduces rotavirus infection among the oldest age group. These indirect effects may be crucial for maintaining the benefits of vaccination in endemic settings. Because adults are more likely to become infected through water sources than children [42], understanding this indirect pathway is especially important in endemic settings.

In small communities like ours, transmission dynamics are strongly dependent on pathogen introduction and outbreaks for some pathogens like rotavirus cannot be maintained in isolation. The importance of pathogen introduction most likely varies by pathogen and depends on the ability for pathogens to persist in environmental sources. We previously showed that the gradient in infection risk by remoteness is strongest for *E. coli*, intermediate for rotavirus, and weakest for *Giardia*, which could be explained by differences in environmental persistence and shedding duration by pathogen [68]. Outside of the diarrheal disease literature, pathogens that can also infect wildlife reservoir species, like Chagas disease [151] might also be able to circulate within relatively isolated communities in the absence of periodic reintroductions, whereas infections like measles that cannot

persist outside of human hosts [152] might also show similar dependence on pathogen introduction [153].

Given that rotavirus can be transmitted through both direct contact and the environment, our results suggest that neither interventions that affect direct transmission alone (such as hand hygiene) nor those that impact water transmission only (such as water treatment) will be sufficient to stop transmission. For this reason, the rotavirus vaccine is an important tool that can greatly enhance rotavirus control because it impacts both transmission pathways. However, more work is needed to understand its effects in community settings and to determine if vaccination will be successful in the long term. Follow up studies could also help determine if one of these two transmission pathways is more strongly impacted than the other.

APPENDICES

APPENDIX A

Appendix for Chapter 2

A.1 Causal Diagram

The causal diagram relating demographic variables to travel is shown in eFigure1 below. Baseline remoteness is considered a common cause of all demographic variables as well as a cause of travel patterns (not shown).

A.2 Transmission Model Details

We developed a deterministic, compartmental susceptible, infected, recovered (SIR) transmission model with two patches at a time. One patch was parameterized to reflect a generic close, medium, or far village, and the other patch had the same size and population structure as Borbón. All disease processes were parameterized to represent rotavirus. Contact between the communities occurred when members of the in-region village traveled to the out-of-region city. Each simulation only considered one village (close, medium, or far) in addition to the city; interactions between the modeled village and other villages were not considered both because these communities villages are relatively small and would

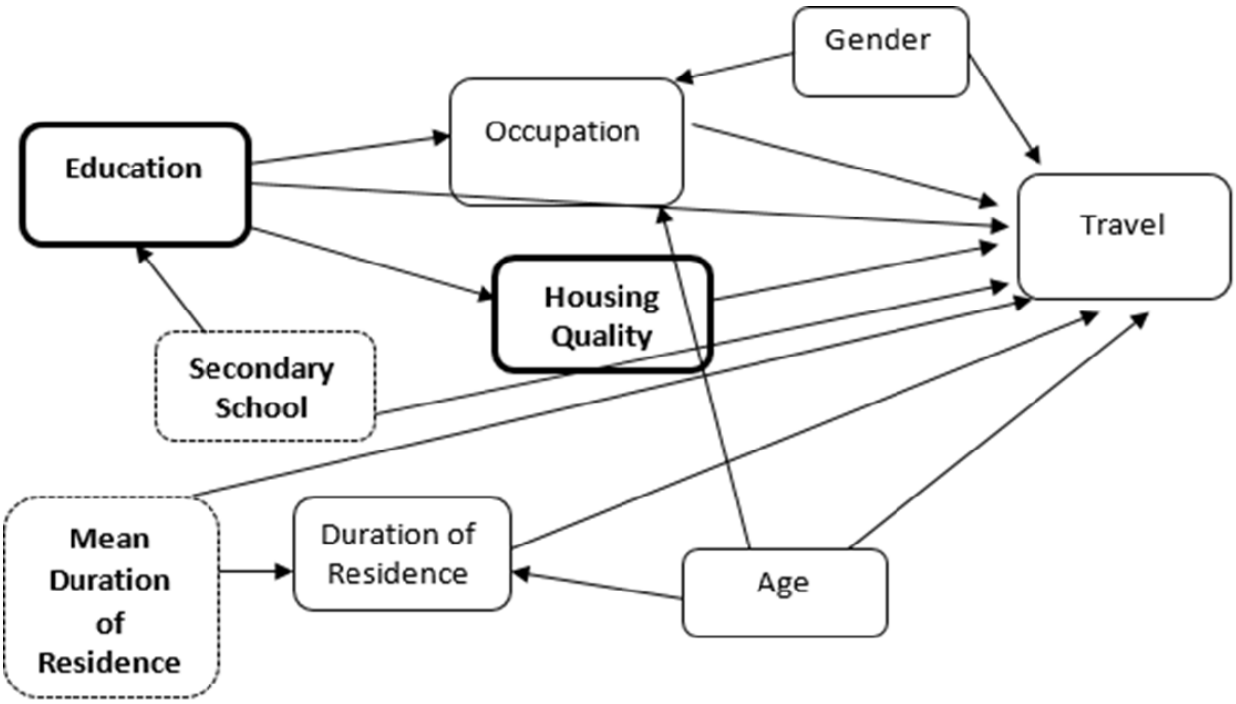

Figure A.1: Demographic Variables Conceptual Diagram. Dashed boxes, thick boxes, and think boxes are shown for community, household, and individual variables respectively.

not be expected to sustain an outbreak in isolation and because of the geography of our study region. Although villagers might leave our study region to travel to other remote communities and not urban centers, due to the geography and transportation network of our study region, these travelers would have to pass through the larger city of Borbon to access transportation to other more distant communities. We stratified the model by age, dichotomized at five years, to capture heterogeneity in risk of infection and travel. Children under five were more likely to become infected and less likely to travel. Both travel patterns and disease transmission parameters were allowed to vary by age. For this model, we allow travelers to infect and be infected during travel. To do this, we allowed members of the village compartments to move back and forth to a corresponding "traveling" compartment imagined to be physically located at the other patch; the transmission rates of villagers in the traveling compartment are the same as the people in that new patch until they return, following the method of Knipl [87]. We also assume that infected and susceptible individuals have the same travel probabilities.

This model includes three different transmission relevant processes: (1) infection, (2) recovery, and (3) travel. We used survey data to estimate parameters for infection and travel. The recovery rate and relative transmission rates for adults and children were taken from prior literature [31]. To estimate the transmission β terms for each community, we used age-specific prevalence of infection data from stool samples collected as part of the population-based case control study in 2007, before the vaccine was introduced. Because our study region is near its endemic equilibrium, we used steady state formulas to calculate the value of R_0 based on life expectancy and average age at first infection [37], and corrected for the fraction of cases that were symptomatic. We used prevalence ratios by remoteness [16] to fix the value of relative β terms by remoteness. To parameterize the travel portion of our model, we used the survey data for each survey year (2003, 2007, 2010, and 2013) to derive daily travel rates to the city for close, medium, and far villages. Additionally, we used census data collected as part of the EcoDess study to account for the

relative population sizes of the different community population sizes and the fraction of the village in each age group for each of the four study years.

In this model, the $S_{i,j}$, $I_{i,j}$, and $R_{i,j}$ compartments represent susceptible, infectious, and recovered individuals in patch i, where $i = 1$ is the within-region village and $i = 2$ is the city, and of age group j (where A is $\leq$5 and C is $<$ 5). Compartments $S^*_{1,j}$, $I^*_{1,j}$, and $R^*_{1,j}$ denote villagers of age j who have traveled to the city. The daily transmission rate from age group j to age group k in community i is given by $\beta_{jk,i}$. The travel rate from community i to ℓ is given by $\tau_{i\ell}$. The recovery rate is given by γ and is independent of age or community. The ordinary differential equations for this model are shown below.

$$
\begin{aligned}
\dot{S}_{1,C} =& - S_{1,C}[\beta_{CC,1} I_{1,C} + \beta_{AC,1} I_{1,A}] - \tau_{12,C} S_{1,C} + \tau_{21,C} S^*_{1,C} \\
\dot{I}_{1,C} =& S_{1,C}[\beta_{CC,1} I_{1,C} + \beta_{AC,1} I_{1,A}] - \tau_{12,C} I_{1,C} + \tau_{21,C} I^*_{1,C} - \gamma I_{1,C} \\
\dot{R}_{1,C} =& \gamma I_{1,C} - \tau_{12,C} R_{1,C} + \tau_{21,C} R^*_{1,C} \\
\dot{S}_{1,A} =& - S_{1,A}[\beta_{AA,1} I_{1,A} + \beta_{CA,1} I_{1,C}] - \tau_{12,A} S_{1,A} + \tau_{21,A} S^*_{1,A} \\
\dot{I}_{1,A} =& S_{1,A}[\beta_{AA,1} I_{1,A} + \beta_{CA,1} I_{1,C}] - \tau_{12,A} I_{1,A} + \tau_{21,A} I^*_{1,A} - \gamma I_{1,A} \\
\dot{R}_{1,A} =& \gamma I_{1,A} - \tau_{12,A} R_{1,A} + \tau_{21,A} R^*_{1,A} \\
\dot{S}^*_{1,C} =& - S^*_{1,C}[\beta_{CC,2} (I^*_{1,C} + I_{2,C}) + \beta_{AC,2} (I^*_{1,A} + I_{2,A})] - \tau_{21,C} S^*_{1,C} + \tau_{12,C} S^*_{1,C} \\
\dot{I}^*_{1,C} =& S^*_{1,C}[\beta_{CC,2} (I^*_{1,C} + I_{2,C}) + \beta_{AC,2} (I^*_{1,A} + I_{2,A})] - \tau_{21,C} I^*_{1,C} + \tau_{12,C} I_{1,C} - \gamma I_{1,C} \\
\dot{R}^*_{1,C} =& \gamma I *_{1,C} - \tau_{21,C} R *_{1,C} + \tau_{12,C} R_{1,C} \\
\dot{S}^*_{1,A} =& - S^*_{1,A}[\beta_{AA,2} (I^*_{1,A} + I_{2,A}) + \beta_{CA,2} (I^*_{1,C} + I_{2,C})] - \tau_{21,A} S^*_{1,A} + \tau_{12,A} S^*_{1,A} \\
\dot{I}^*_{1,A} =& S^*_{1,A}[\beta_{AA,2} (I^*_{1,A} + I_{2,A}) + \beta_{CA,2} (I^*_{1,C} + I_{2,C})] - \tau_{21,A} I^*_{1,A} + \tau_{12,A} I_{1,A} - \gamma I_{1,A} \\
\dot{R}^*_{1,A} =& \gamma I^*_{1,A} - \tau_{21,A} R^*_{1,A} + \tau_{12,A} R_{1,A} \\
\dot{S}_{2,C} =& - S_{2,C}[\beta_{CC,2} (I_{2,C} + I^*_{1,C}) + \beta_{AC,2} (I^*_{1,A} + I_{2,A})] \\
\dot{I}_{2,C} =& S_{2,C}[\beta_{CC,2} (I_{2,C} + I^*_{1,C}) + \beta_{AC,2} (I^*_{1,A} + I_{2,A})] - \gamma I_{2,C} \\
\dot{R}_{2,C} =& \gamma I_{2,C} \\
\dot{S}_{2,A} =& - S_{2,A}[\beta_{AA,2} (I_{2,A} + I^*_{1,A}) + \beta_{CA,2} (I^*_{1,C} + I_{2,C})] \\
\dot{I}_{2,A} =& S_{2,A}[\beta_{AA,2} (I_{2,A} + I^*_{1,A}) + \beta_{CA,2} (I^*_{1,C} + I_{2,C})] - \gamma I_{2,A} \\
\dot{R}_{2,A} =& \gamma I_{2,A}
\end{aligned}
$$

To fully understand the impact of travel and transmissibility on disease risk (measured by cumulative incidence for all models), we used a series of models. To assess the overall changes over time and their effects on risk, we used survey data from all time points in

our study region for close, medium, and far villages to estimate the net effects of these on risk of rotavirus. The derivation of all parameters and their values are described in detail below. However, because many variables changed for each time point simultaneously, we also conducted what we call a pure effects analysis where we changed only the travel parameters and kept all other factors constant. For our first pure effects model, we systematically increased the average travel. For the second pure effects model, we fixed the travel patterns of children and increased the travel of adults by a proportionality constant to explore the role of heterogeneity. This model allowed us to assess the effect of heterogeneity in travel on disease risk. Using these models, we were able to assess to what extent the changes in our study region were driven by demographic change versus changes in travel.

A.3 Model Parameterization

We used a combination of literature and survey data to parameterize our model. Parameter values and their sources are shown in eTable 1 below. When the source is "Estimated," we describe the estimation in a later section.

Table A.1: Parameters used in model simulations. All parameters have units of days^{-1}.

Parameter	Value				Source
	Close	Medium	Far	City	
Child-Child Transmission Rate (β_{CC})	0.275	0.5	0.331	0.66	Estimated
Child-Adult Transmission Rate (β_{CA})	0.138	0.25	0.166	0.33	Estimated
Adult-Child Transmission Rate (β_{AC})	0.275	0.5	0.331	0.66	Estimated
Adult-Adult Transmission Rate (β_{AA})	0.138	0.25	0.166	0.33	Estimated
Village-city Travel Rate for Adults ($\tau_{12,A}$)	0.029	0.027	0.014	—	Estimated
City-Village Travel Rate for Adults ($\tau_{21,A}$)	1	1	1	—	Estimated
Village-city Travel Rate for Children ($\tau_{12,C}$)	0.017	0.03	0.0063	—	Estimated
City-Village Travel Rate for Children ($\tau_{21,C}$)	1	1	1	—	Estimated
Recovery Rate (γ)	0.2	0.2	0.2	0.2	[3]

In addition to differences in disease and travel parameters by community, we accounted for their different population sizes by fixing the relative proportions of susceptibles in the out of region community and the community of interest. To get the size of the 'close',

'medium' and 'far' villages, we used the mean population size for all time points for the 15 communities in our study, stratified by remoteness. We elected to use the mean for all time points and not the last time point because there were no significant time trends in population size. The mean population sizes were 575, 108, and 246 for close, medium, and far villages respectively. We assume that the out of region village has a population size of 5,000, comparable to Borbón.

A.3.1 Transmission Rate Calculations

Because our study region is close to its endemic equilibrium, we used the following steady state formula to calculate the value of R_0 [91]:

$$R_0 = L/A$$

In this equation, L is the average life expectancy and A is the average age of first infection. Other researchers have used an alternative formulation of R_0 ($R_0 = 1+\frac{L}{A}$) but later analysis by Dietz has suggested that omitting the 1 ($R_0 = \frac{L}{A}$) is more accurate [91]. We also conducted sensitivity analysis for a wide range of β values and these results are presented in a later section. We assumed a life expectancy of 70 years and approximated the average age of first infection in our study region using the average age of those infected in 2007, corrected for the fraction symptomatic. In 2007, the average age of first infection was 15.7 for all cases and 11.4 for symptomatic cases. Since the only pathway of infection in this model is direct transmission, the average R_0 is simply $\frac{\beta}{\gamma}$. Fixing the value of γ to 0.2, we estimated that the average β value for the region overall was 0.315. To estimate the local transmission parameters for close, medium, and far villages, we used ratios of

prevalence of infection by remoteness as a proxy for the β term of that village [68]. In this paper, the prevalence of symptomatic infection was found to be 0.6, 0.9, and 0.5 per 100 persons for remote, medium, and close villages respectively and 1.2 per 100 persons in Borbón. Taking Borbón as the reference, these values translate to prevalence ratios of 0.5, 0.75, and 0.42 for remote, medium and close villages versus Borbón respectively. To be consistent with prior literature, we assumed that transmission to children had a β term that was twice as high as transmission to adults [3]. Thus the estimated β term for a given age group within a particular village was a weighted average of the transmission patterns of adults and children based on their population size.

Patch Specific R_0

For the full model, the absolute value of R_0 is not especially informative. What is more useful is the patch-specific R_0 values, which suggest the transmission potential within a community after the outbreak is seeded by travel to the city. The reduced model for a single village and the resulting R_0 calculations are shown below.

$$\begin{aligned}
\dot{S}_A =& - S_A[\beta_A I_A + \beta_{CA} I_C] \\
\dot{I}_A =& S_A[\beta_A I_A + \beta_{CA} I_C] - \gamma I_A \\
\dot{R}_A =& \gamma I_A
\end{aligned}$$

$$\begin{aligned}\dot{S}_C =& - S_C[\beta_C I_C + \beta_{AC} I_A] \\ \dot{I}_C =& S_C[\beta_C I_C + \beta_{AC} I_A] - \gamma I_C \\ \dot{R}_C =& \gamma I_C\end{aligned}$$

Using the next generation matrix approach, we calculate the matrices F, V, and V^{-1}. The largest eigenvalue of the matrix product FV^{-1} is R_0. The calculations that lead to this result are shown below.

$$f = \begin{bmatrix} \beta_A S_A I_A + \beta_{CA} S_A I_C \\ \beta_C S_C I_C + \beta_{AC} S_C I_A \end{bmatrix}$$

$$v = \begin{bmatrix} \gamma I_A \\ \gamma I_C \end{bmatrix}$$

Now we take the Jacobians of both matrices at the disease free equilibrium to produce F and V.

$$\mathcal{F} = \begin{bmatrix} \beta_A S_A & \beta_{CA} S_A \\ \beta_{AC} S_C & \beta_C S_C \end{bmatrix}$$

$$\mathcal{V} = \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}$$

Then, we take the inverse of V

$$V^{-1} = \begin{bmatrix} 1/\gamma & 0 \\ 0 & 1/\gamma \end{bmatrix}$$

Then the matrix product FV^{-1} is

$$FV^{-1} = \begin{bmatrix} \frac{\beta_A S_A}{\gamma} & \frac{\beta_{CA} S_A}{\gamma} \\ \frac{\beta_{AC} S_C}{\gamma} & \frac{\beta_C S_C}{\gamma} \end{bmatrix}$$

The largest Eigenvalue of this matrix was calculated using Mathematica and is:

$$\frac{\sqrt{(\beta_A)^2 (S_A)^2 - 2\beta_A \beta_{AC} S_A S_C + (\beta_{AC})^2 (S_C)^2} + \beta_A S_A + \beta_{AC} S_C}{2\gamma}$$

A.3.1.1 Global Reproduction Number

In all cases, the system R_0 once travel is included are greater than 1 because of travel to the city.

Although the basic reproduction number for the full model will always be greater than 1 in our study villages, we show our basic setup to allow other researchers to apply our findings to other settings in which travel is needed to bring the global R_0 above 1.

To calculate R_0, we begin by calculating the matrices F and V representing the rates of new

infections and net compartment transfer, respectively.

$$f = \begin{bmatrix} S_{1,C}\left(\beta_{CC,1} I_{1,C} + \beta_{AC,1} I_{1,A}\right) \\ S_{1,A}\left(\beta_{CA,1} I_{1,C} + \beta_{AA,1} I_{1,A}\right) \\ S^*_{1,C}\left(\beta_{CC,2}(I_{2,C} + I^*_{1,C}) + \beta_{AC,2}(I_{2,A} + I^*_{1,A})\right) \\ S^*_{1,A}\left(\beta_{CA,2}(I_{2,C} + I^*_{1,C}) + \beta_{AA,2}(I_{2,A} + I^*_{1,A})\right) \\ S_{2,C}\left(\beta_{CC,2}(I_{2,C} + I^*_{1,C}) + \beta_{AC,2}(I_{2,A} + I^*_{1,A})\right) \\ S_{2,A}\left(\beta_{AC,2}(I_{2,C} + I^*_{1,C}) + \beta_{AA,2}(I_{2,A} + I^*_{1,A})\right) \end{bmatrix}$$

$$v = \begin{bmatrix} \tau_{12,C} I_{1,C} - \tau_{21,C} I^*_{1,C} + \gamma I_{1,C} \\ \tau_{12,A} I_{1,A} - \tau_{21,A} I^*_{1,A} + \gamma I_{1,A} \\ \tau_{12,C} I^*_{1,C} - \tau_{21,C} I_{1,C} + \gamma I^*_{1,C} \\ \tau_{12,A} I^*_{1,A} - \tau_{21,A} I_{1,A} + \gamma I^*_{1,A} \\ \gamma I_{2,C} \\ \gamma I_{2,A} \end{bmatrix}$$

Now we take the Jacobians of both matrices at the disease free equilibrium

$$\mathcal{F} = \begin{bmatrix} \beta_{CC,1}\bar{S}_{1,C} & \beta_{AC,1}\bar{S}_{1,C} & 0 & 0 & 0 & 0 \\ \beta_{CA,1}\bar{S}_{1,A} & \beta_{AA,1}\bar{S}_{1,A} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{CC,2}\bar{S}^*_{1,C} & \beta_{AC,2}\bar{S}^*_{1,C} & \beta_{CC,2}\bar{S}^*_{1,C} & \beta_{AC,2}\bar{S}^*_{1,C} \\ 0 & 0 & \beta_{CA,2}\bar{S}^*_{1,A} & \beta_{AA,2}\bar{S}^*_{1,A} & \beta_{CA,2}\bar{S}^*_{1,A} & \beta_{AA,2}\bar{S}^*_{1,A} \\ 0 & 0 & \beta_{CC,2}\bar{S}_{2,C} & \beta_{AC,2}\bar{S}_{2,C} & \beta_{CC,2}\bar{S}_{2,C} & \beta_{AC,2}\bar{S}_{2,C} \\ 0 & 0 & \beta_{CA,2}\bar{S}_{2,A} & \beta_{AA,2}\bar{S}_{2,A} & \beta_{CA,2}\bar{S}_{2,A} & \beta_{AA,2}\bar{S}_{2,A} \end{bmatrix}$$

$$\mathcal{V} = \begin{bmatrix} \tau_{12,C}+\gamma & 0 & -\tau_{21,C} & 0 & 0 & 0 \\ 0 & \tau_{12,A}+\gamma & 0 & -\tau_{21,A} & 0 & 0 \\ -\tau_{12,C} & 0 & \tau_{21,C}+\gamma & 0 & 0 & 0 \\ 0 & -\tau_{12,A} & 0 & \tau_{21,A}+\gamma & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma \end{bmatrix}$$

We then take the inverse of V.

$$V^{-1} = \begin{bmatrix} \frac{\gamma+\tau_{21,C}}{\gamma^2+2\tau_{12,C}\tau_{21,C}+\gamma(\tau_{12,C}+\tau_{21,C})} & 0 & -\frac{\tau_{21,C}}{\gamma^2+2\tau_{12,C}\tau_{21,C}+\gamma(\tau_{12,C}+\tau_{21,C})} & 0 & 0 & 0 \\ 0 & \frac{\gamma+\tau_{21,A}}{\gamma(\gamma+\tau_{12,A}+\tau_{21,A})} & 0 & \frac{\tau_{21,A}}{\gamma(\gamma+\tau_{12,A}+\tau_{21,A})} & 0 & 0 \\ \frac{\tau_{12,C}}{\gamma^2+2\tau_{12,C}\tau_{21,C}+\gamma(\tau_{12,C}+\tau_{21,C})} & 0 & \frac{\gamma+\tau_{12,C}}{\gamma^2+2\tau_{12,C}\tau_{21,C}+\gamma(\tau_{12,C}+\tau_{21,C})} & 0 & 0 & 0 \\ 0 & \frac{\tau_{12,A}}{\gamma(\gamma+\tau_{12,A}+\tau_{21,A})} & 0 & \frac{\gamma+\tau_{12,A}}{\gamma(\gamma+\tau_{12,A}+\tau_{21,A})} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\gamma} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\gamma} \end{bmatrix}$$

To calculate R_0, the next steps are to calculate the matrix product FV^{-1} and its eigenvalues. The largest eigenvalue of the next generation matrix FV^{-1} is the global R_0. While this matrix product and its eigenvalues can be calculated, the next generation matrix is

quite large and the resulting eigenvalues are on the order of 6000 terms long. For this reason, they are not presented here. Readers interested in calculating the global R_0 for similar systems might wish to do so numerically.

A.3.2 Travel Rate Calculations

The fraction of people in a given age group traveling to the out-of-region city in the last seven days is known from survey data and is denoted ω. Assume that the time between travel is exponentially distributed with mean $1/\tau$. Then, the relationship between ω and τ is

$$\omega = \int_0^7 \tau e^{-\tau t}\, dt = 1 - e^{-7\tau},$$

implying

$$\tau = -\frac{\log(1-\omega)}{7}.$$

For example, if 17.5% of people (the value for close villages) traveled in the last seven days, then τ =0.0275, with an average of travel every 36 days. Similarly, if only 8.5% of people (the value for far villages) traveled in the last seven days, then $\tau = 0.013$, with an average of travel every 79 days. We assume duration of travel is short, one day on average.

A.4 Simulation Approach

For each model, we simulated for 200 days with an infection seed of 1% of the population in the city and calculated the cumulative incidence for each age group as well as the fraction of infections acquired in the in-region village.

We investigated the impact of changing the average frequency of travel, the average rate of transmission, and heterogeneity in both factors by age (children vs. adults). To investigate the average effects of travel and transmission, we used the range of travel and infectivity parameters estimated from our data. For travel, we used the measured values of travel for each remoteness level. Because the village-specific R_0 values in our study region varied from 0.79 (close villages) to 1.43 (medium villages), we used the β parameters from these two village types to represent the range of transmission rates seen in our study region. To investigate heterogeneity in travel, we kept the travel rates of children constant at their measured value for close villages (τ=.017 visits/day) and increased travel rates for adults only by a proportionality constant. To investigate heterogeneity in transmission, we fixed the transmission rates of adults and increased transmission for children by a proportionality constant.

A.5 Net Effects: Parameter Values

The parameter values for each study year for the net effects models are shown in the tables below. Parameters for close, medium, and far villages are shown in eTable 2, eTable 3, and eTable 4 respectively. Parameters for the city are shown in eTable 5 below. The parameters that were varied for the pure effects models are also indicated with a † symbol in each table.

A.6 Sensitivity Analysis

To enhance the generalizability of our findings to other regions that might have different levels of disease transmissibility and population structures, we also conducted sensitivity analysis across a range of β values and population sizes. We also considered a model with three age groups to account for the greater complexity of our data ($<$5, 5–13, $>$13) but

Table A.2: Model parameters for close villages for each study year. Population sizes were based on census data. Travel was based on the proportion reporting travel in the sociometric survey, which was then used to derive the specific τ parameters, as described above. We used stool sample data to estimate transmission parameters but assumed the ratio children: adults was 2 and fixed gamma based on prior literature [3]

	2003	2007	2010	2013
Population size (Survey Data)				
Overall	616	683	741	686
Proportion $<$ 5 years	0.192	0.221	0.205	0.145
Proportion $>=$ 5 years	0.808	0.779	0.795	0.855
Ratio Population Village: City	0.11	0.12	0.13	0.12
Travel				
Survey Data				
Overall Travel	0.082	0.14	0.075	0.175
Travel $<$ 5	0.043	0.081	0.069	0.113
Travel $>=$ 5	0.093	0.156	0.077	0.185
Derived Model Parameters				
τ Average †	0.012	0.022	0.011	0.027
$\tau_{12,C}$	0.006	0.012	0.010	0.017
$\tau_{12,A}$	0.014	0.024	0.011	0.029
τ Ratio (Adult to Child)†	2.22	2.01	1.12	1.71
Transmission Parameters				
β_C	0.275	0.275	0.275	0.275
β_{AC}	0.275	0.275	0.275	0.275
β_A	0.138	0.138	0.138	0.138
β_{CA}	0.138	0.138	0.138	0.138
γ	0.2	0.2	0.2	0.2
Within Patch R_0	0.82	0.84	0.83	0.79

†Varied for pure effects models

Table A.3: Model parameters for medium villages for each study year. Population sizes were based on census data. Travel was based on the proportion reporting travel in the sociometric survey, which was then used to derive the specific τ parameters, as described above. We used stool sample data to estimate transmission parameters but assumed the ratio children: adults was 2 and fixed gamma based on prior literature [3]

	2003	2007	2010	2013
Population size (Survey Data)				
Overall	119	118	121	123
Proportion $<$ 5 years	0.214	0.221	0.196	0.135
Proportion $>=$ 5 years	0.786	0.779	0.804	0.865
Ratio Population Village: City	0.023	0.023	0.024	0.024
Travel				
Survey Data				
Overall Travel	0.071	0.08	0.092	0.155
Travel $<$ 5	0.015	0.013	0.044	0.189
Travel $>=$ 5	0.086	0.099	0.104	0.15
Derived Model Parameters				
τ Average†	0.011	0.012	0.014	0.024
$\tau_{12,C}$	0.002	0.002	0.006	0.030
$\tau_{12,A}$	0.013	0.015	0.016	0.023
τ Ratio (Adult to Child) †	5.95	7.97	2.44	0.776
Transmission Parameters				
β_C	0.5	0.5	0.5	0.5
β_{AC}	0.5	0.5	0.5	0.5
β_A	0.25	0.25	0.25	0.25
β_{CA}	0.25	0.25	0.25	0.25
γ	0.2	0.2	0.2	0.2
Within Patch R_0	1.52	1.53	1.5	1.42

† Varied for pure effects models

Table A.4: Model parameters for far villages for each study year. Population sizes were based on census data. Travel was based on the proportion reporting travel in the sociometric survey, which was then used to derive the specific τ parameters, as described above. We used stool sample data to estimate transmission parameters but assumed the ratio children: adults was 2 and fixed gamma based on prior literature [3]

	2003	2007	2010	2013
Population size (Survey Data)				
Overall	307	305	326	313
Proportion $<$ 5 years	0.196	0.197	0.187	0.141
Proportion $>=$ 5 years	0.804	0.803	0.813	0.859
Ratio Population Village: City	0.058	0.057	0.061	0.059
Travel				
Survey Data				
Overall Travel	0.04	0.063	0.054	0.086
Travel $<$ 5	0.021	0.023	0.034	0.043
Travel $>=$ 5	0.044	0.072	0.058	0.093
Derived Model Parameters				
τ Average†	0.006	0.009	0.008	0.013
$\tau_{12,C}$	0.003	0.003	0.005	0.006
$\tau_{12,A}$	0.006	0.011	0.009	0.014
τ Ratio (Adult to Child)	2.12	3.21	1.73	2.22
Transmission Parameters				
β_C	0.331	0.331	0.331	0.331
β_{AC}	0.331	0.331	0.331	0.331
β_A	0.166	0.166	0.166	0.166
β_{CA}	0.166	0.166	0.166	0.166
γ	0.2	0.2	0.2	0.2
Within Patch R_0	0.99	0.99	0.99	0.95

† Varied for pure effects models

Table A.5: Model parameters for city for each study year. Population sizes were assumed constant and based on Borbon and the ratio of children to adults was set to like close villages. We did not consider travel from the city except returning villagers. Since we assumed the travel duration was 1 day, all τ values for the city were fixed to 1. As for the villages, we used stool sample data to estimate transmission parameters but assumed the ratio children:adults was 2 and fixed gamma based on prior literature [3]

	2003	2007	2010	2013
Population size				
Overall	5000	5000	5000	5000
Proportion $<$ 5 years	0.192	0.221	0.205	0.145
Proportion $>=$ 5 years	0.808	0.779	0.795	0.855
Travel Parameters				
τ Average †	1	1	1	1
$\tau_{21,C}$	1	1	1	1
$\tau_{21,A}$	1	1	1	1
τ Ratio (Adult to Child)†	1	1	1	1
Transmission Parameters				
β_C	0.66	0.66	0.66	0.66
β_{AC}	0.66	0.66	0.66	0.66
β_A	0.33	0.33	0.33	0.33
β_{CA}	0.33	0.33	0.33	0.33
γ	0.2	0.2	0.2	0.2
Within Patch R_0	1.97	2.01	1.99	1.89

†Varied for pure effects models for the villages, but not the city

found that the results were qualitatively similar.

In all of our analyses, the city was the major driver of transmission. To determine if this effect was predominantly driven by the population size of the city relative to the village or the relative infectivity in the city, we systematically varied both characteristics.

A.6.1 Population Size

For this sensitivity analysis, travel patterns for adults and children were fixed to the average for close villages (0.027) for both children and adults. The transmission terms for the village (patch 1) were allowed to differ by age and were set to the estimated values for medium villages. As stated in the text, as the population sizes of the city and village became more similar, the fraction of transmission attributable to local transmission increased proportionally with population size but the overall incidence decreased. See eTable 6 below.

Table A.6: Cumulative incidence and proportion of locally acquired infection by relative population size.

Village population size (% of total population)	Low within-village transmission		High within-village transmission	
	Children	Adults	Children	Adults
Cumulative Incidence				
5%	0.056	0.029	0.057	0.030
10%	0.052	0.027	0.055	0.028
15%	0.048	0.025	0.053	0.027
20%	0.044	0.022	0.050	0.025
25%	0.038	0.019	0.046	0.023
Proportion Locally Acquired				
5%	0.03	0.03	0.06	0.06
10%	0.06	0.06	0.12	0.11
15%	0.10	0.09	0.17	0.17
20%	0.13	0.13	0.24	0.23
25%	0.16	0.16	0.30	0.30

A.6.2 Transmissibility

We also considered the impact of varying the transmissibility of rotavirus in each community. We considered three approaches: 1) increasing the transmissibility (β terms) in the village only, such that the local R_0 of the village and the city became more similar; 2) increasing transmissibility for both the village and the city proportionally, such that the ratio of the R_0 for the village and the city remained the same; and 3) allowing transmission parameters to be equal (such that the ratio of R_0 values was 1) in the city and the village but increasing the value of R_0. This increase in R_0 was done by keeping transmission to adults constant and increasing it for children only.

In scenario 1, we found that increasing the R_0 of the village had no effect on risk (see eTable 7), but impacted the distribution of cases for children and led to more local transmission. When both the transmissibility of the village and the city were increased (scenario 2, see eTable 8), the total number of cases increased. The fraction of cases that were locally acquired increased slightly, but the majority of cases were acquired in the city. When the values of R_0 for the city and village were kept equal, the city remained the primary source of cases although the total number of cases increased as the value of R_0 increased.

Table A.7: Sensitivity analysis for scenario 1. The R_0 of the city is constant and the R_0 of the village increases.

R_0 Ratio	City R_0	Village R_0	Fraction City Children	Fraction City Adults	Relative Infectivity (Children vs. Adults)	Cumulative Incidence (CI)
2.74	1.89	0.69	0.92	0.93	1	0.04
2.39	1.89	0.79	0.87	0.92	2	0.04
2.12	1.89	0.89	0.82	0.92	3	0.04
1.91	1.89	0.99	0.77	0.92	4	0.04
1.73	1.89	1.09	0.72	0.92	5	0.04
1.59	1.89	1.19	0.68	0.92	6	0.04
1.47	1.89	1.29	0.65	0.92	7	0.04
1.36	1.89	1.39	0.61	0.92	8	0.04
1.27	1.89	1.49	0.58	0.92	9	0.04
1.19	1.89	1.59	0.55	0.91	10	0.04

For all three scenarios, the majority of cases originated in the city, regardless of the local

Table A.8: Sensitivity analysis for scenario 2. The R_0 and the village and the city are both increased but their ratio is kept constant

R_0 Ratio	City R_0	Village R_0	Fraction City Children	Fraction City Adults	Relative Infectivity (Children vs. Adults)	Cumulative Incidence (CI)
2.39	1.65	0.69	0.88	0.92	1	0.03
2.39	1.89	0.79	0.87	0.92	2	0.04
2.39	2.13	0.89	0.86	0.92	3	0.04
2.39	2.37	0.99	0.85	0.91	4	0.05
2.39	2.61	1.09	0.84	0.91	5	0.05
2.39	2.85	1.19	0.83	0.91	6	0.05
2.39	3.08	1.29	0.83	0.90	7	0.05
2.39	3.32	1.39	0.81	0.89	8	0.06
2.39	3.56	1.49	0.80	0.89	9	0.06
2.39	3.80	1.59	0.79	0.89	10	0.06

Table A.9: Sensitivity analysis for scenario 3. The transmission parameters for the village and the city are equal but the value of R_0 is increased.

R_0 Ratio	City/Village R_0	Fraction City Children	Fraction City Adults	Relative Infectivity (Children vs. Adults)	Cumulative Incidence (CI)
1.00	0.69	0.96	0.97	1	<0.01
1.00	0.79	0.95	0.97	2	<0.01
1.00	0.89	0.95	0.97	3	<0.01
1.00	0.99	0.94	0.97	4	<0.01
1.00	1.09	0.94	0.96	5	<0.01
1.00	1.19	0.94	0.96	6	<0.01
1.00	1.29	0.93	0.96	7	0.01
1.00	1.39	0.93	0.96	8	0.01
1.00	1.49	0.93	0.96	9	0.01
1.00	1.59	0.92	0.95	10	0.01

transmission parameters. Even when the village β term for children was over 3.5 times that of children in the city, the majority of village child cases still originated in the city. We therefore concluded that the major reason for the influence of the city was its population size.

APPENDIX B

Appendix for Chapter 3

B.1 Additional Descriptive Graphics for Meta-Analysis

The distribution of temperature from the studies included in the meta-analysis is shown in Figure B.1. The majority of studies were in the $<$10°C or 20-30 °C range, with very few studies conducted in temperatures over 35°C.

The distribution of the estimated die-off rates from all 39 studies is shown in Figure B.2. The distribution was right skewed-many studies had decay rates that were close to 0 with a long tail of studies showing higher decay.

The relationship between temperature and decay rate by the type of water in which the experiment was conducted is shown in Figure B.3. The studies conducted in natural waters appeared to have the highest decay rates. The distribution of experimental temperatures by water type was uneven, with higher proportions of natural water samples tested at higher temperatures. No obvious interaction was observable between temperature and water type, which is unsurprising given the low number of studies conducted in treated, combined, or purified water at higher temperatures.

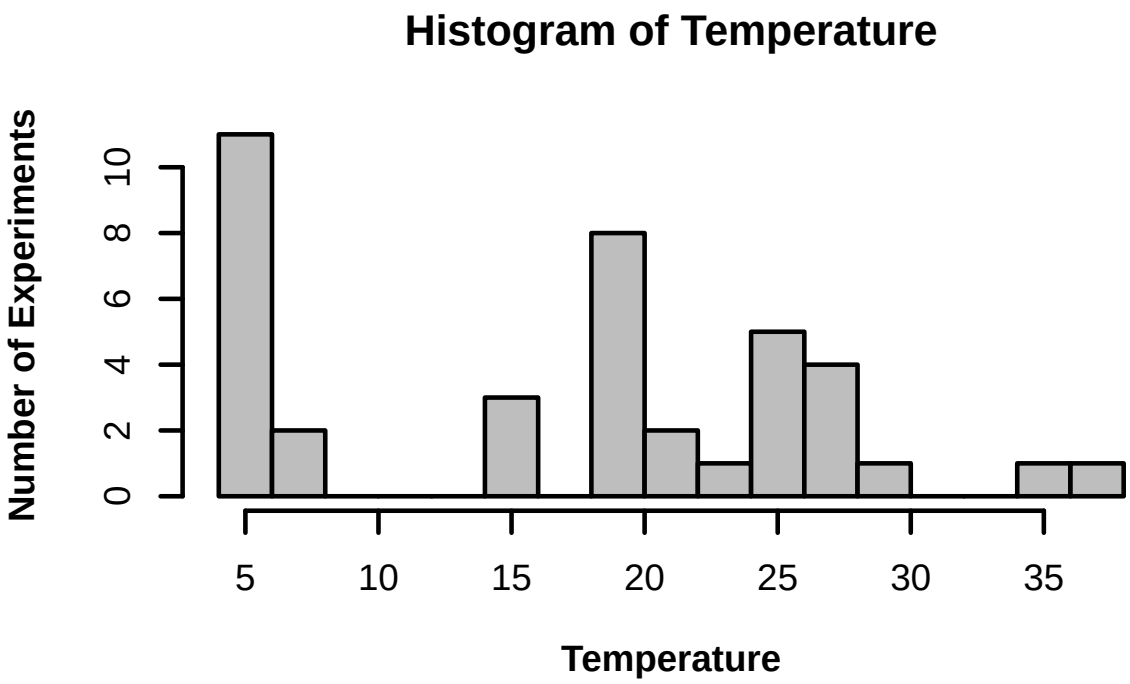

Figure B.1: Distribution of temperature from studies included in the meta-analysis

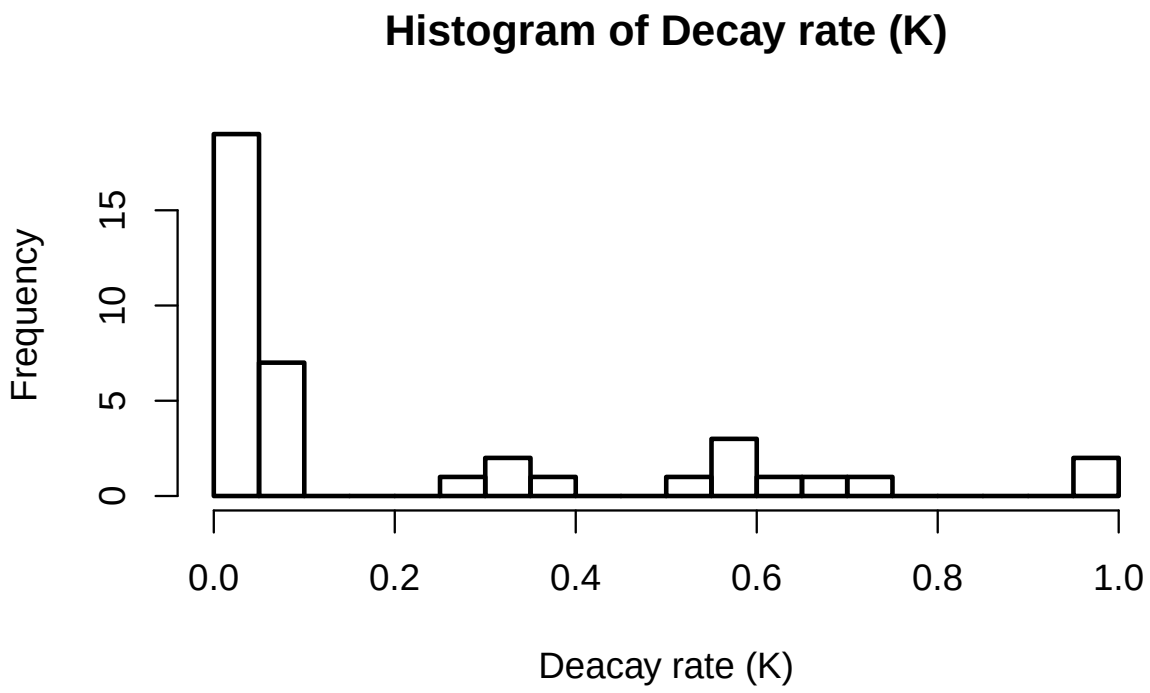

Figure B.2: Distribution of decay rates estimated from studies included in the meta-analysis

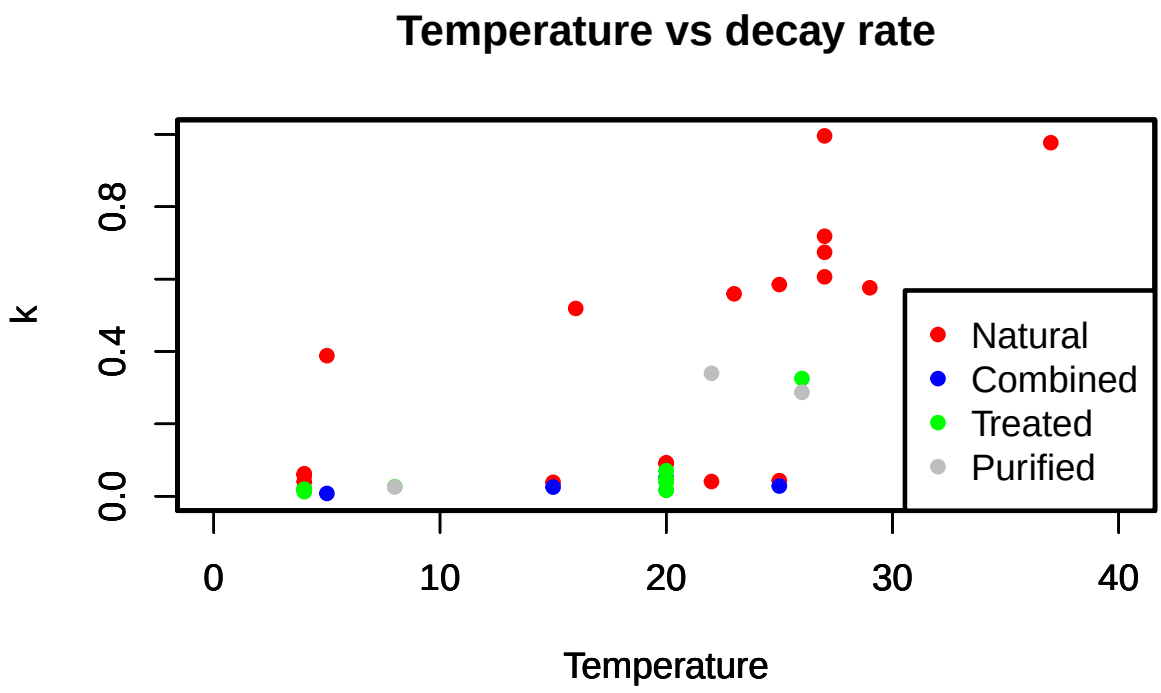

Figure B.3: Relationship between temperature and decay rate by water type

The relationship between temperature and decay rate by study is shown in Figure B.4. While only one study considered temperatures over 30°C, the estimated association between temperature and decay rates was generally consistent across studies. Thus the temperature effect was not completely explained by a study effect.

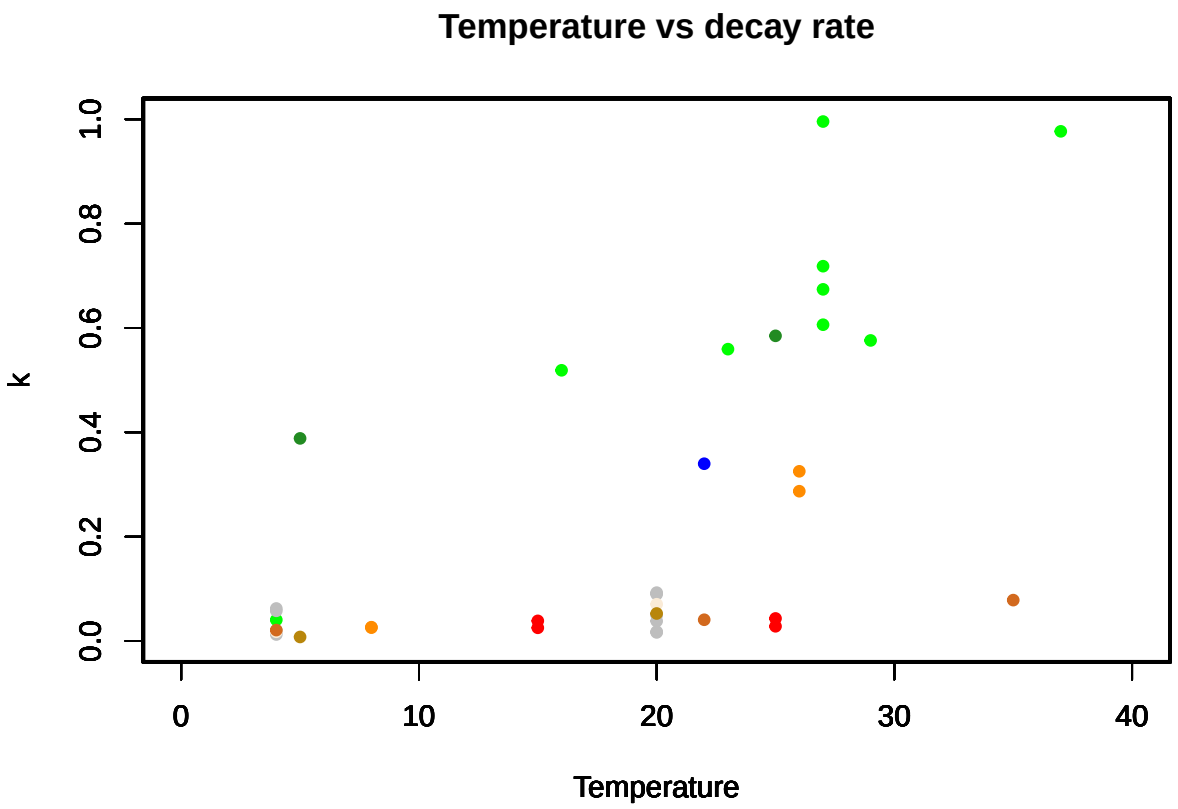

Figure B.4: Temperature versus decay rate by study (different studies shown in different colors). In general, decay rates were positively associated with temperature for all studies.

B.2 Multivariate prediction of decay rates

Because the studies were conducted in different types of water solutions, we also estimated the effect of water type on decay to determine if the temperature association was confounded by differences in water type. We expected the primary reason for differences

in rotavirus decay rates between different water types to be antiviral activity of microbial communities, with natural water sources being the only source likely to contain other microbes or their virucidal secretions. For this reason, and because we had few observations for all of the different water types, we used a binary indicator for water type categorizing studies into natural or non-natural water classes. We also tested to see if the effect of temperature was modified by this water type indicator, which would indicate that temperature has a stronger effect on decay in certain types of water sources. Results of the regression modeling are shown in Table B.1 below.

Table B.1: Regression results. The unadjusted model contained only one variable at a time: temperature or water type. The adjusted model contained both temperature and water type. The adjusted + interaction model contained temperature, the natural water sources indicator, and their interaction

	Unadjusted estimate (s.e.)	**Adjusted** estimate (s.e.)	**Adjusted + Interaction** estimate(s.e.)
Temperature	0.095 (0.019)	0.075(0.016)	0.070(0.021)
Natural Water	1.62(0.39)	1.21(0.33)	1.03(0.44)
Temperature*Natural Water			-0.02(0.03)

Although water type was also a significant predictor of decay rate, the effect of temperature remained significant. Furthermore, the interaction between water type and temperature was not significant, indicating that our data did not contain evidence for variable effects of temperature by water type. We elected to use the results from the adjusted model to parameterize the transmission model. The results were qualitatively similar using the unadjusted model. The maximum temperature effect was 3% per degree Celsius for the unadjusted model compared with 2.4% for the adjusted model. A plot of the temperature effect for both models is shown in Figure B.5.

Although we fit this model on the log scale, we have also plotted the results on the linear scale to illustrate the predicted decay rates by temperature (Figure B.6).

We acknowledge that there may be some residual confounding by water type (and other

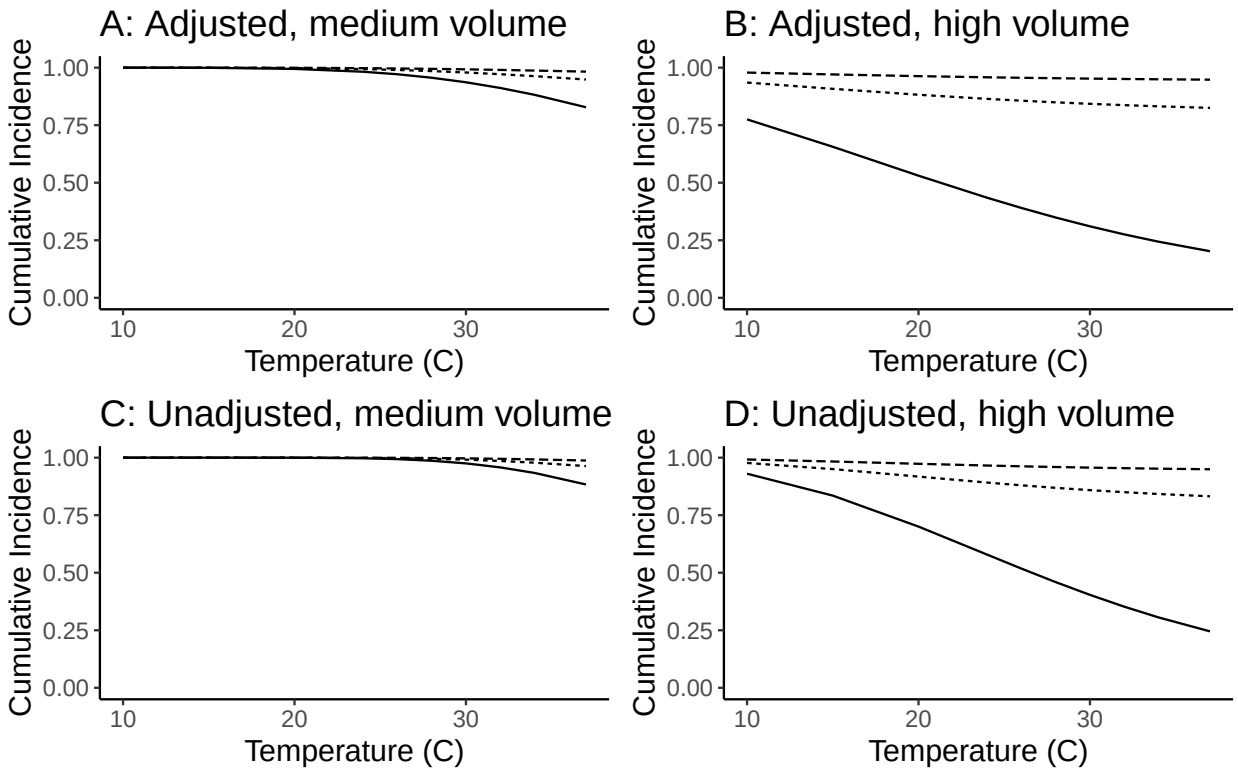

Figure B.5: Cumulative incidence of rotavirus infection (fraction) as a function of temperature for different levels of direct transmission (dashed: $\mathcal{R}_{0,H} = 3$; dotted: $\mathcal{R}_{0,H} = 2$; and solid: $\mathcal{R}_{0,H} = 1$). Adjusted estimates come from the adjusted model in table S1, which was adjusted for water type (natural water sources=1, other=0). Estimates come from standing water with A) moderate volume (1,000 L/person, 10^6 L total volume) from the adjusted model B) Moderately high volume (10,000 L/person, 10^7 L total volume) and the adjusted model, C) moderate volume from the unadjusted model, and D) moderately high volume for the unadjusted model. The maximum effect of temperature occurs at higher volumes because at lower volumes $\mathcal{R}_{0,W}$ is greater than 1 for all temperatures.

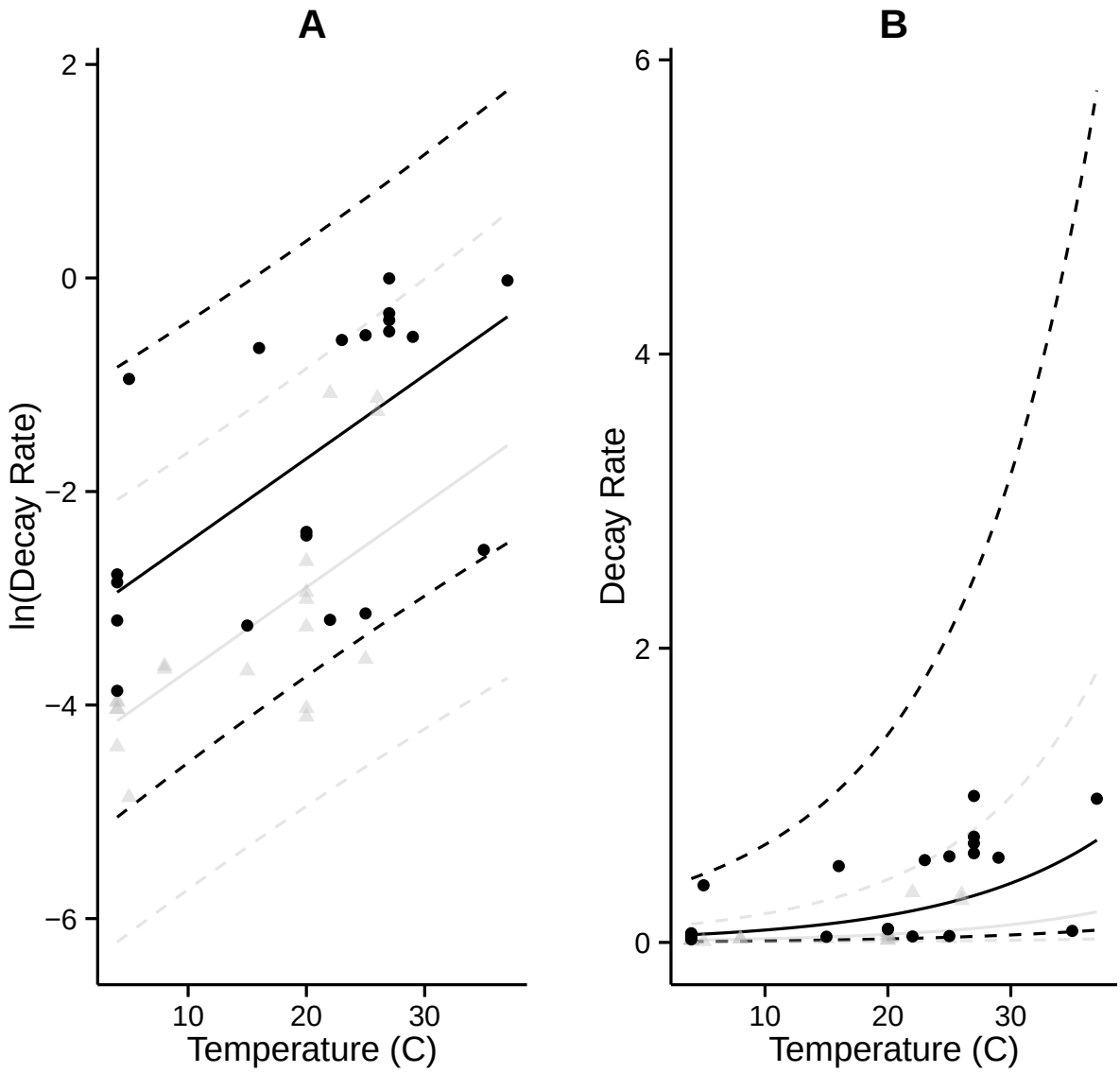

Figure B.6: Predicted relationship between temperature and decay rates on the log scale (panel A) and the linear scale (panel B).

variables) and that there may also be other confounders or effect modifiers that impact this relationship. We were underpowered to detect these factors here, but future studies investigating the impact of other environmental variables in addition to temperature would be useful to help address this concern.

B.3 The mixing cell model

B.3.1 Overall Structure

Most water quality models parameterize pathogen transport using the advection-diffusion equation (ADE). The ADE describes the movement of a 'packet' (i.e. the mass input at one location during a single time step) of solutes or suspended constituents in flow as a distributed concentration, with mean position proportional to average downstream flow velocity (advection) and variance in position (due to fluctuations in the velocity field, i.e. from turbulence) proportional to a diffusion or dispersion rate constant. The ADE partial differential equation in one dimension is:

$$\frac{\partial W}{\partial t} = -v\frac{\partial W}{\partial x} + D\frac{\partial^2 W}{\partial x^2} + G$$

Where W is pathogen concentration within some frame of reference, is time, is flow velocity, is downstream distance, is the rate of diffusion or dispersion, and is a term for sources or sinks of pathogens, such as shedding of pathogens to the waterway, deposition / resuspension processes or decay. Under the ADE, at time t the distribution of a packet pathogens along the flow axis, referenced to the source location, is normal with mean vt and variance 2Dt.

A commonly used discrete distance approximation to the ADE represents river reaches as networks of interconnected cells, and the change of pathogen concentration over time for each cell is represented with a biased random walk. In a non-branching network, this formulation may be written as:

$$\frac{dW_i}{dt} = \frac{L_i}{v_b}(PW_{i-1} + (1-P)W_{i+1} - (P + (1-P))W_i) + G$$

Where W_i is pathogen concentration in cell i of the network, L_i/v_b (the length of cell divided by a velocity term) is the average residence time of pathogens in cell i, and $P[0,1]$ is the bias term, or probability that a pathogen moves downstream upon leaving the cell. The net downstream advective velocity is given by the difference in average downstream rate of movement and average upstream rate of movement, i.e. $v_b(P-(1-P)) = v_b(2P-1)$. For a series of cells of equal length L, the variance in position of pathogens at time t is v_bLt. Thus, the biased random walk formulation approximates the ADE when v_b, P, and L are chosen such that $v = v_b(2P-1)$ and $D = v_bL/2$.

The mixing cell formulation used in this paper is a further simplification, equivalent to the biased random walk model when $P = 1$:

$$\frac{dW_i}{dt} = \frac{L_i}{v}(W_{i-1} - W_i)$$

In the mixing cell model, no upstream movement is possible, and pathogen concentrations become Poisson distributed with mean position (relative to initial position) vt and variance vLt. Thus, while behavior at the upstream border of pathogen sources differs somewhat

from the ADE, the two models behave similarly in flowing systems when $D = vL/2$. Cumulative pathogen residence times per length of stream reach, the key determinant of epidemiological consequences of hydrological transport, behave similarly in the ADE, biased random walk, and mixing cell formulations. The simpler biased random walk and mixing cell implementations have advantages with respect to numerical tractability and stability when implemented on larger scales.

B.3.2 The use of intermediate boxes

We found that using only one box overestimates the probability of a pathogen surviving downstream. A proof is shown below.

In each box, pathogens die at rate μ and leave for the next box at rate $np\nu$ where n is the number of boxes the length is divided into. If there is just one box, the probability that the pathogen survives downstream is

$$\frac{p\nu}{p\nu + \mu}.$$

In reality, we should find the probability as $n \to \infty$,

$$\lim_{n\to\infty} \left(\frac{np\nu}{np\nu + \mu} \right)^n \tag{B.1}$$

$$= \lim_{n\to\infty} \left(\frac{np\nu + \mu}{np\nu} \right)^{-n} \tag{B.2}$$

$$= \lim_{n\to\infty} \left(1 + \left(\frac{\mu}{p\nu} \right) \frac{1}{n} \right)^{-n} \tag{B.3}$$

$$= \exp\left(-\frac{\mu}{p\nu} \right). \tag{B.4}$$

This gives the true probability of a pathogen surviving downstream.

Now, let $x = \mu/(p\nu)$. In the one-box approximation (Bertuzzo, etc), they estimate the

survival of the pathogen to be $1/(1+x)$ while the true probability is e^{-x}. Since

$$\frac{1}{1+x} > e^{-x},$$

the approximation will overestimate the likelihood of an outbreak downstream.

Through simulation, we found that only one box is needed to correct this problem. However, we elected to use 10 boxes in our model to make the unit conversions more transparent. To determine the number of boxes needed, we used three models. These models are diagramed in Figure B.7 below. In all three models, we model 10 villages that are each 10 km apart, volume is low (100 L/person), and the effective river velocity is 10 km/day. Results were similar for other initial conditions.

In model 1, the villages are contiguous and no intermediate boxes are used. In this example, river speed is 10 km/d, so the rate of leaving a 10km long water box is 10/10=1/d. In model 2, there are no people outside of the 2km long borders of the villages. There is one intermediate water box. River speed is 10km/d, so the rate of leaving a 2km long water box is 10/2=5/d, and the rate of leaving an 8km long water box is 10/8= 1.25/d. In model 3, there are no people outside the 2 km borders of the village and there are 4 intermediate boxes (each corresponding to a 2 km length of river). Since river speed is 10 km/d, the rate of leaving a 2 km river box is 10/2=5/day. We simulated each of these three models for 150 days and the resulting incidence curves are shown in Figure B.8 below.

B.3.3 Unit conversions

B.3.3.1 River Velocity

To convert measured flow velocities to daily rates with units of inverse days, we divided the measured flow velocity by the length of river represented by each compartment in our

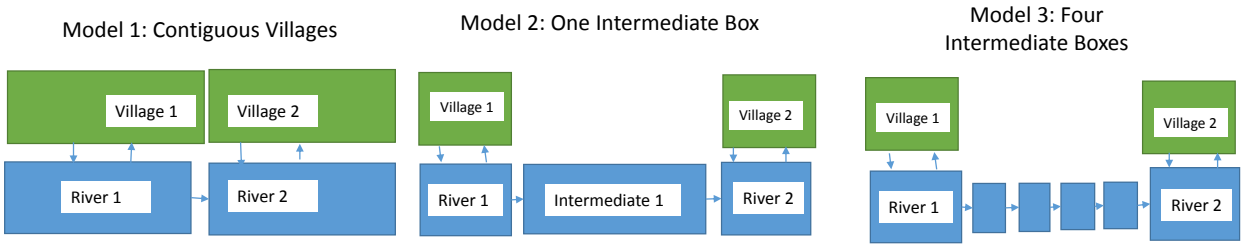

Figure B.7: Conceptual diagram of the three models used to evaluate the number of boxes needed.

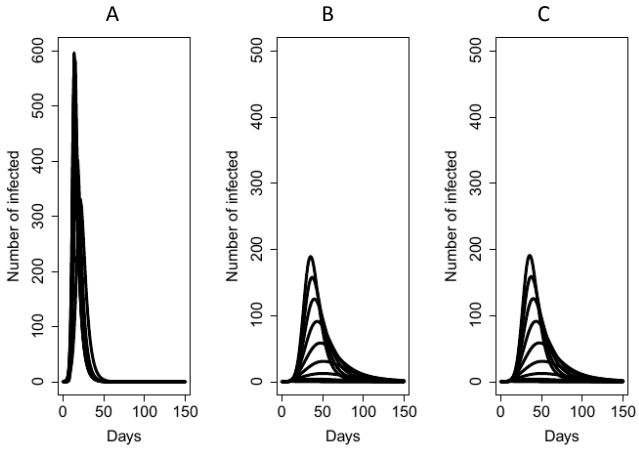

Figure B.8: Incidence curves for each of the three models. Models 1, 2, and 3 are shown in figure panels A, B, and C respectively.

model (called l, measured in km) as below:

$$v(/day) = \frac{Velocity(km/day)}{l(km)}$$

Because we chose to have each compartment in our model represent a 1 km stretch of river, the measured flow velocities and the parameter v have the same numerical values, but different units.

In our distributed delay model, since each box is equal to a 1 km stretch of river, v was simply equal to the overall measured river flow velocity, in km/day. Since there was no upstream flow in this model and the first village was located at the source of the river, no pathogens flowed into community 1. The water supply for community 1 is assumed to arrive via runoff and the volume of the water supply in both communities is determined by recent rainfall.

B.3.3.2 Dose response curve

The force of infection in an environmentally mediated infectious disease transmission model is

$$\beta_W \rho \pi W$$

Where β is the contact rate, ρ is the volume of water per contact, π is the probability of infection per pathogen consumed, and W is the concentration of pathogens. Now, we want

to scale W by the median infectious dose, N_{50}, such that:

$$W^* = \frac{W}{N_{50}}$$

Because N_{50} is the median infectious dose, we have

$$\pi N_{50} = 0.5$$

So, in terms of W*, the force of infection is

$$\beta_W \rho \pi W = \beta_W \rho \pi (N_{50} W^*) = 0.5(\beta_W \rho W^*)$$

B.4 Calculating the volume of the water reservoir

B.4.1 Average Volume

We based our volume calculations on work by Downing et al. that describes the size distributions of lakes worldwide [109]. In this analysis, the minimum surface area analyzed was 0.001 km and the average surface area of lakes worldwide was 0.012 km^2. Because we conceptualized the water reservoir as a section of a flowing stream, we would not expect the volume of water accessible to a community along the water front at any one time to be greater than that of an average lake. Given this size, we selected values of α to roughly correspond to a lake with a surface area of 0.012 km^2 (the average surface area found by [109]) and a depth of 10 m (maximum α). The moderate volume scenario roughly

corresponds to this same surface area with a depth of 1 m. The low volume scenario was more similar to the volume expected for a small pond.

B.4.2 Effect of Seasonal Rain

In our study region, the water level for the Cayapas increases by an average of 0.3 m between December and January, when the rainy season generally begins. For the Santiago River, this increase in rainfall is about 0.48 m. Assuming channel width is fixed, this increase in depth corresponds to a volume increases of 3200 and 4800 L for the Cayapas and Santiago rivers respectively. To assess the effect of changes in seasonal rain on temperature effects, we fixed the baseline volume and then increased rainfall by 3200 L for the Cayapas river and 4800 L for the Santiago River. The result of this sensitivity analysis is shown in Figure B.9 below. The increased volume resulting from seasonal differences in water level was not expected to dilute the reservoir enough to lead to a decrease in risk. Therefore, it is likely that most of the changes in risk due to rainfall occur on much shorter timescales or more sporadically, e.g., during flooding events which temporarily raise the river level much more dramatically. Such transient effects of dilution are beyond the scope of this paper and are not considered here.

B.5 Calculating the basic reproduction number

To calculate the basic reproduction number, we use the model equations corresponding to the source of the river network, where the outbreak is seeded. These equations are shown below.

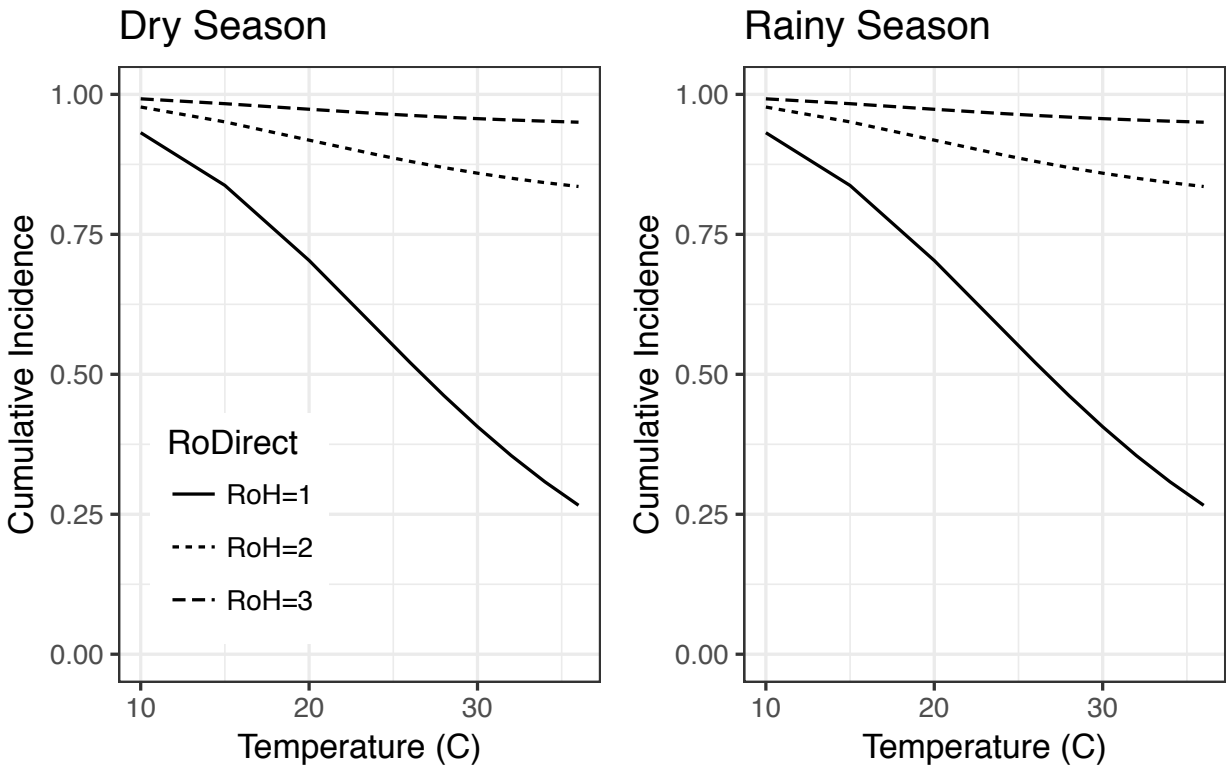

Figure B.9: Relationship between temperature and cumulative incidence by level of seasonal rain. Panels A, B, and C respectively show these relationships when the strength of direct transmission is 1, 2, and 3 respectively. The color of the points indicates the temperature used in the simulation with the lowest temperature (darker color) producing the highest risk and the higher temperatures being shown in lighter color.

$$
\begin{aligned}
\dot{S} =& -\frac{\beta_H SI}{N} - \frac{\beta_W S\rho W}{2} \\
\dot{I} =& \frac{\beta_H SI}{N} + \frac{\beta_W S\rho W}{2} - \gamma I \\
\dot{R} =& \gamma I \\
\dot{W} =& \frac{(1-c)\phi}{V} I - \mu W - vW
\end{aligned}
$$

Using the next generation matrix approach, we calculate the matrices F, V, and V^{-1}. The largest eigenvalue of the matrix product FV^{-1} is R_0. The calculations that lead to this result are shown below. For ease of calculation, we replace $(1-c)\phi$ with the quantity ϕ*. In these calculations, we adapt the notation ϕ for ϕ*.

First, we define vectors corresponding to the generation of new cases (the f vector) and all other terms (the v vector). These vectors are shown below. There are two infected compartments in our model (I and W), so there are two items in each vector.

$$
f = \begin{bmatrix} \frac{\beta_H SI}{N} + \frac{\beta_W S\rho W}{2} \\ 0 \end{bmatrix}
$$

$$
v = \begin{bmatrix} \gamma I \\ \mu W + vpW - \frac{\phi *}{V} \end{bmatrix}
$$

Now we take the Jacobians of both vectors to produce 2 by 2 matrices. The matrix F is the matrix corresponding to the generation of new cases in each compartment and V

corresponds to the net transfer out of each compartment.

$$\mathcal{F} = \begin{bmatrix} \frac{\beta_H S}{N} & \frac{\beta_W S \rho}{2} \\ 0 & 0 \end{bmatrix}$$

$$\mathcal{V} = \begin{bmatrix} \gamma & 0 \\ -\phi * /V & \mu + v \end{bmatrix}$$

At the disease free equilibrium, V does not change because it does not depend on the relative number of people in each compartment, but F becomes:

$$\mathcal{F} = \begin{bmatrix} \beta_H & \frac{\beta_W N \rho}{2} \\ 0 & 0 \end{bmatrix}$$

Then, we take the inverse of V

$$V^{-1} = \begin{bmatrix} \frac{1}{\gamma} & 0 \\ \frac{\phi * /V}{\gamma(v+\mu)} & \frac{1}{v+\mu} \end{bmatrix}$$

Then the matrix product FV^{-1} is

$$FV^{-1} = \begin{bmatrix} \frac{\beta_W}{\gamma} + \frac{0.5\beta_W N \rho \phi *}{V\gamma(v+\mu)} & \frac{0.5\beta_W N \rho}{v+\mu} \\ 0 & 0 \end{bmatrix}$$

The largest Eigenvalue of this matrix is:

$$\begin{aligned} R_0 =& \frac{\beta_H}{\gamma} + \frac{0.5\beta_W N\rho\frac{\phi*}{V}}{\gamma(v+\mu)} \\ =& R_{0,H} + R_{0,W} \end{aligned}$$

Substituting for $\phi* = (1-c)\phi$, the R_0 equation can be re-written as follows:

$$R_0 = \frac{\beta_H}{\gamma} + \frac{0.5\beta_W N\rho\frac{(1-c)\phi}{V}}{\gamma(v+\mu)}$$

For cities, the volume of the water reservoir (V) is better interpreted as a per capita water availability. In this case, we can replace V with αN, where the parameter α describes the per capita water availability. The equation then becomes:

$$R_0 = \frac{\beta_H}{\gamma} + \frac{0.5\beta_W \rho(1-c)\phi}{\alpha\gamma(v+\mu)}$$

APPENDIX C

Appendix for Chapter 4

C.1 Defining vaccine effectiveness parameters

We use the counterfactual framework described by Halloran and Hudgens [140] to estimate vaccine effectiveness. For all equations, we let two doses of vaccine be defined by a parameter called $vaccinated$, which takes a value of 1 for fully vaccinated individuals and a value of 0 for unvaccinated individuals. The variables α and α' define the regional or community coverage of vaccination. Each variable is defined in its corresponding section. In all equations, $\bar{y}$ is a statistic that quantifies the average outcome (rate or odds depending on the model) for a given covariate pattern. All effectiveness parameters are conceptually defined based on a counterfactual contrast, but all variables in these equations are represented by statistics within our sample because all individuals at risk are not observed (i.e., our statistics are random variables that depend on our sampling process).

Vaccine effectiveness measures can be classified into four groups: direct effects, indirect effects, overall effects, and total effects. A diagram of each of these possible effectiveness measures and how they were estimated in our study is shown in Figure C.1.

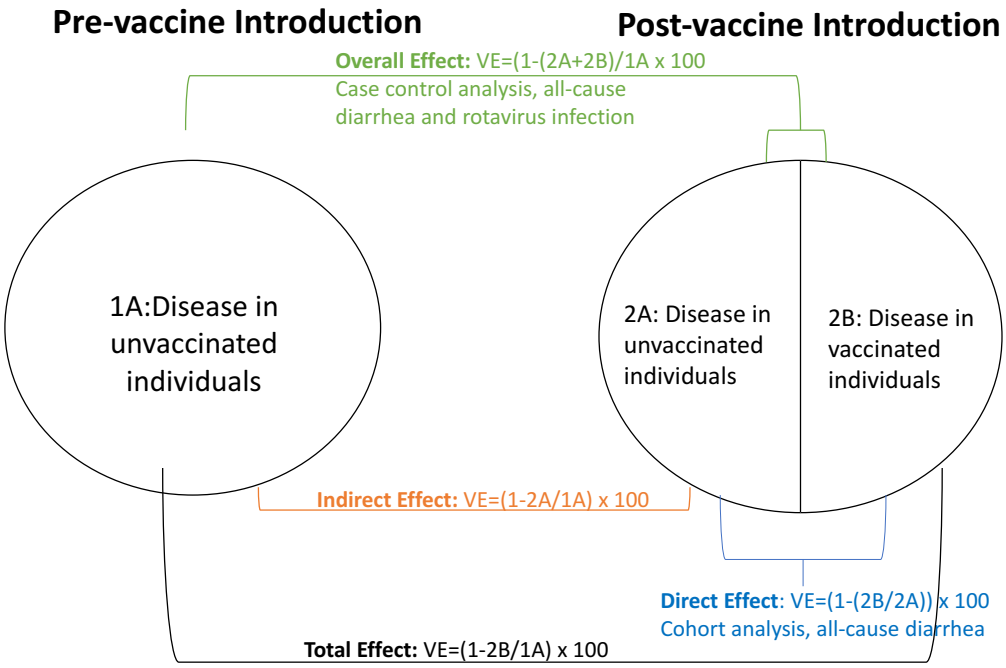

Figure C.1: Diagram showing how each vaccine effectiveness measure was calculated. This figure was adapted from Panozzo et al, 2014 [2]. In this analysis, we estimated the overall effect of vaccination using data from the case control study (green) and the direct effect using data from the diarrheal disease surveillance to conduct a cohort analysis (blue). While not directly estimated, the overall effect of vaccination is equivalent to the indirect effect (orange) for older children and adults (age ≥ 5) because there were no vaccinated individuals in that age group (i.e., 2B=0). The total effect (black) was also not directly estimated, but is approximated by the overall effect for younger age groups when vaccine coverage is high (i.e., 2A is approximately equal to 0).

C.1.1 Direct Effect (DE)

The general equation for calculating direct effects is shown below.

$$\overline{DE}(\alpha) = \bar{y}(vaccinated = 0|\alpha) - \bar{y}(vaccinated = 1|\alpha)$$

Under 5 years of age In our analysis, we estimate the direct effect of vaccination on all cause diarrhea using Poisson regression analysis. To estimate this quantity, we compare the rate of diarrhea for fully vaccinated children with children who were not vaccinated. We estimate this quantity for all study communities simultaneously, so α is equal to the average vaccine coverage after introduction (around 76%). We estimate this quantity separately for 2 doses and 1 dose and code the model as below using dummy variables. In the equation below, the 'offset' term quantifies the person-time observed for each individual. The intercept term, β_0 is the estimated rate of diarrhea when all other covariates (X_i) are equal to 0. Then, each β_{X_i} term is the multiplicative increase in the rate of diarrhea for a one unit increase in covariate X_i for continuous variables or an indicator variable for it's presence/absence for binary variables. The variable ε_i is the error not captured by our regression model.

$$\log(\mu_i) = \log(\text{offset}) + \beta_0 + \beta_{2doses} X_{i,2doses} + \beta_{1dose} X_{i,1dose} + ... + \varepsilon_i$$

Because this equation estimates the average rate of rotavirus infection for a given covariate pattern, the difference in outcomes is given by exponentiating β_{2dose} for two doses

of vaccine and β_{1dose} for one dose of vaccine. When the interaction term with age>2 years is added to the model, this contrast gives the vaccine efficacy for children under 2 years of age. So the indirect effect of vaccination is given by $(1 - \exp(\beta_{2doses}) \times 100 = (1-\text{IRR}_{2dose}) \times 100$ for two doses of vaccine and $(1-\exp(\beta_{1dose}) \times 100 = (1-\text{IRR}_{1dose}) \times 100$. This equation takes the same form for non-diarrheal illness.

Over 5 years of age Because no one over the age is vaccinated with rotavirus, there is no direct effect of vaccination.

C.1.2 Indirect Effect (IE)

The general equation for calculating indirect effects is shown below.

$$\overline{IE}(\alpha, \alpha') = \bar{y}(vaccinated = 0; \alpha) - \bar{y}(vaccinated = 0; \alpha')$$

Under 5 years of age Because we do not know the vaccination status of every child after the vaccine was introduced and because only 8 children who received 0 doses of vaccine were sampled in the case control study after vaccine introduction, we are not able to estimate the indirect effect on this age group.

Over 5 years of age Because no one over the age of five is vaccinated with rotavirus, only the indirect effect of vaccination can be calculated for this age group and in this case the overall effect is equal to the indirect effect. In our analysis, we estimate this indirect using logistic regression analysis for the sub-sample of our case control population over 5 years of age.

We compare the effect of vaccine introduction in general, so that α takes a value of 0 and α'

is equal to the average vaccine coverage after introduction. To compare rotavirus infection and diarrheal illness before and after the vaccine was introduced, we can code a dummy variable for cycle which takes values of 0 before the vaccine was introduced and during the first cycle of introduction and takes a value of 1 beginning 1 study cycle after the vaccine is introduced to account for the fact that children must seroconvert before adults can be indirectly protected.

The logistic regression model for rotavirus infection and for diarrhea illness in the case control then has the following form:

$$\log\left(\frac{P(Rota=1)}{1-P(Rota=1)}\right) = \beta_0 + \beta_{introduced} X_{i,introduced} + ... + \varepsilon_i$$

In this equation, the intercept term, β_0 is the estimated odds of rotavirus infection when all other covariates (X_i) are equal to 0. Then, each β_{X_i} term is the multiplicative increase in the odds for a one unit increase in covariate X_i for continuous variables or an indicator variable for it's presence/absence for binary variables. The variable ε_i is the error not captured by our regression model.

As described in the method section, this regression analysis was weighted to account for complex sampling design. Because this equation estimates the average odds of rotavirus infection for a given covariate pattern, the difference in outcomes is given by exponentiating $\beta_{introduced}$. So the indirect effect of vaccination is given by $(1 - \exp(\beta_{vaccinated})) \times 100 = (1 - OR) \times 100$. This equation takes the same form for diarrheal infection.

C.1.3 Total Effect (TE)

The general equation for calculating total effects is shown below.

$$\overline{TE}(\alpha, \alpha') = \bar{y}(vaccinated = 0; \alpha) - \bar{y}(vaccinated = 1; \alpha')$$

Under 5 years of age. Unfortunately, we were not able to directly estimate this quantity with confidence using the data that we had.

Ideally, this quantity would be calculated by comparing the rates of diarrhea for children with two doses of vaccine after introduction with the general unvaccinated population before vaccination (i.e., comparing $\alpha = 0$ with $\alpha' = 0.76$). However, because we only had case control data for 53 children who were fully vaccinated and 3 children who were vaccinated with one dose of vaccine, we could not assess the effectiveness of vaccination on rotavirus infection.

To estimate this quantity for all cause diarrhea, We could have compared the rates of diarrhea in the 2011-2013 surveillance for fully vaccinated individuals with all children under five in the 2003-2007 surveillance but we elected not to do this because our missing data analysis suggested that children who did not have vaccine records were not comparable to those with vaccine records in 2011-2013. As such, in order to estimate this quantity with minimal bias, we would need to compare this two dose group with children who received as least one vaccine from 2003-2007. These earlier vaccine records had been lost, so we were not able to make this comparison. To deal with this issue, we could have matched children with vaccine data to other children during the surveillance period by education and community (accounting for the key differences between children with and without

vaccine records in the 2011-2013 surveillance), but such an analysis would have required assuming that the two time periods were otherwise comparable (i.e., there were no secular trends not attributable to vaccination). We previously showed that the strains of rotavirus circulating in this region have changed rapidly over time and that strains varied widely between seasons in our case control analysis. For this reason, we were not comfortable making this assumption.

However, the total effect should be similar to the overall effect of vaccination for children less than 5 years of age when vaccine coverage is high (estimation procedure for overall effects is described below).

Over 5 years of age. Because individuals over five are not vaccinated, this quantity is not defined for this group.

C.1.4 Overall Effect (OE)

The general equation for calculating overall effects is shown below.

$$\overline{OE}(\alpha, \alpha') = \bar{y}(\alpha) - \bar{y}(\alpha')$$

For all groups, we estimated the overall effects using case control data on rotavirus positivity. We considered both vaccine introduction (binary) and vaccine coverage (continuous with values between 0 and 1) for each age group separately and for the population overall. We found that actual vaccine coverage predicted odds of rotavirus infection among children and the general population in later cycles, after the vaccine had been introduced. This procedure is described more in detail below.

Under 5 years of age We compared both the effect of vaccine introduction in general using a dummy variable as described above (so that $\alpha = 0$ and $\alpha' = 0.76$) and using the vaccine coverage among children with vaccine records, treating alpha as a continuous variable (so that $\alpha = 0$ and $\alpha' = 1$).

For vaccine introduction, this gives:

$$\log\left(\frac{P(Rota = 1)}{1 - P(Rota = 1)}\right) = \beta_0 + \beta_{introduced} X_{i,introduced} + ... + \varepsilon_i$$

For vaccine coverage, this gives:

$$\log\left(\frac{P(Rota = 1)}{1 - P(Rota = 1)}\right) = \beta_0 + \beta_{coverage} X_{i,coverage} + ... + \varepsilon_i$$

When vaccine coverage is high, the contrast of the $\beta_{coverage}$ term describes the comparison between 0 vaccine coverage and 100% vaccine coverage, in which case all eligible individuals are vaccinated so that the overall effect in this group approximates the total effect.

Over 5 years of age. We compared both the effect of vaccine introduction in general using a dummy variable as described above (so that $\alpha = 0$ and $\alpha' = 0.76$) and using the vaccine coverage among children with vaccine records, treating alpha as a continuous variable (so that $\alpha = 0$ and $\alpha' = 1$) but found that because older children and adults were not vaccinated, the actual coverage of the vaccine did not impact the odds of rotavirus after introduction–the entire effect was driven by whether or not the vaccine had been introduced. Thus, we used the binary classification (introduced vs. not introduced) for this

group.

General Population. For the general population, we used the vaccine coverage classification to estimate overall effects because actual coverage was found to be important.

C.1.5 Summary

A summary of the vaccine effectiveness measures calculated in this study and their sources is summarized in the table below.

C.2 Study cycle dates

The dates corresponding to the start of each case control cycle are shown in table C.2.

C.3 Vaccine Data

C.3.1 Data collection details

We collected vaccine data from both local health posts and the regional health center. The regional health center stored copies of all vaccine records from health posts throughout the region, so theoretically these two sources should be redundant. However, in practice there were communities that were in one source and not the other. By triangulating data from both sources, we obtained the maximum amount of data possible. Vaccine records were available for 22 communities from at least one source: six communities had data from the health post only, four communities had data from the regional health centers only, and twelve had data from both sources.

Table C.1: Source of vaccination effectiveness measures

	Direct effect (DE)	Indirect Effect (IE)	Overall Effect (OE)
Age <5			
Rotavirus Infection	–	–	$(1 - exp(\beta_{coverage})) \times 100$ **
All-cause diarrhea	$(1 - exp(\beta_{2dose})) \times 100$*	–	–
	$(1 - exp(\beta_{1dose})) \times 100$ *		
Non-diarrheal illness	$(1 - exp(\beta_{2dose})) \times 100$*	–	–
	$(1 - exp(\beta_{1dose})) \times 100$ *		
Age ≥5			
Rotavirus Infection	N/A	$(1 - exp(\beta_{introduced})) \times 100$**	Equal to IE
General Population			
Rotavirus Infection	N/A	–	$(1 - exp(\beta_{coverage})) \times 100$**

*Estimated using surveillance data and poisson regression

** Estimated using case control data and weighted logistic regression

Table C.2: Cycle dates

Cycle	Date Start	Date End
1	8/01/2003	1/21/2004
2	1/21/2004	10/24/2004
3	11/07/2004	7/28/2005
4	8/01/2005	3/31/2006
5	5/01/2006	12/20/2006
6	1/19/2007	7/24/2007
7	8/10/2007	10/31/2008
8	11/01/2008	12/30/2009
9	1/01/2010	12/31/2010
10	1/01/2011	5/30/2012
11	6/01/2012	12/31/2013

C.3.2 Vaccine coverage

The fraction of community members with vaccine records is shown in table C.3 and the coverage of two doses of vaccine among children with records by community is shown in table C.4. For both tables, communities that were included in the rotavirus positivity analysis are shown in bold. Of children who received both doses of vaccine, most received both doses by 6 months of age (Table C.5).

C.4 Supplemental information for overall effects analysis (part 1, case control analysis)

C.4.1 Population level trends in rotavirus infection and diarrhea over time

The prevalence of diarrhea declined over time for both rotavirus and non-rotavirus diarrhea (Figure C.2). For rotavirus, this change was stronger for asymptomatic rotavirus infection than symptomatic rotavirus infection(Figure C.3).

Table C.3: Vaccine record availability by community. Each cell is presented as Percent with Vaccine Records (Number with Vaccine Records/Number Eligible). Communities that were included in the rotavirus positivity analysis are shown in bold

Community	Cycle 8	Cycle 9	Cycle 10	Cycle 11
Close Communities				
1	**58.3 (7/12)**	**45.0 (9/20)**	**44.4 (8/18)**	**81.8 (9/11)**
2	**43.8 (7/16)**	**61.9 (13/21)**	**34.3 (12/35)**	**0 (0/22)**
3	**44.4 (12/27)**	**67.8 (25/37)**	**80.0 (28/35)**	**85.0 (17/20)**
4	0 (0/7)	45.5 (5/11)	50.0 (8/16)	0 (0/14)
5	30.0 (3/10)	40.0 (4/10)	44.4 (8/18)	0 (0/12)
Total	40.2 (29/72)	56.6 (56/99)	52.5 (64/122)	32.9(26/79)
Medium Communities				
6	50.0 (1/2)	100 (2/2)	50.0 (2/4)	100 (1/1)
7	**66.7 (2/3)**	**54.5 (6/11)**	**100 (6/6)**	**60.0 (3/5)**
8	100 (1/1)	100 (3/3)	40.0 (2/5)	0 (0/7)
9	50.0 (1/2)	80.0 (4/5)	66.7 (2/3)	50.0 (1/2)
10	100 (1/1)	66.7 (2/3)	50.0 (2/4)	100 (1/1)
Total	66.7 (6/9)	70.8 (17/24)	63.6 (14/22)	37.5 (6/16)
Far Communities				
11	**19 (3/16)**	**36.4 (4/11)**	**29.4 (5/17)**	**50.0 (3/6)**
12	100 (7/7)	100 (6/6)	66.7 (2/3)	100 (4/4)
13	0 (0/1)	33.3 (1/3)	75.0 (3/4)	33.3 (1/3)
14	**62.5 (5/8)**	**60.0 (15/25)**	**45.0 (9/20)**	**46.2 (6/13)**
15	75.0 (3/4)	0 (0/3)	0 (0/3)	0 (0/8)
16	0 (0/3)	12.5 (1/8)	40.0 (4/10)	71.4 (5/7)
17	76.9 (10/13)	66.7 (12/18)	50.0 (6/12)	75.0 (12/16)
18	0 (0/2)	40.0 (2/5)	50.0 (2/4)	60.0 (3/5)
19	60.0 (3/5)	50.0 (3/6)	25.0 (5/20)	46.2 (6/13)
20	0 (0/4)	0 (0/3)	0 (0/7)	36.4 (4/11)
21	0 (0/12)	18.2 (2/11)	30.0 (6/20)	72.7 (8/11)
Total	41.3 (31/75)	47.4 (46/97)	35.0 (42/120)	53.6(52/97)

Table C.4: Coverage of two doses of rotavirus vaccine among children with vaccine records. All cells show % Vaccinated (n vaccinated/n with vaccine records). A value of 'N/A' indicates that no vaccine records were available for that cycle.

Community	Cycle 8	Cycle 9	Cycle 10	Cycle 11
Close Communities				
1	**85.7 (6/7)**	**77.8 (7/9)**	**87.5 (7/8)**	**88.9 (8/9)**
2	**57.1 (4/7)**	**46.2 (6/13)**	**50.0 (6/12)**	**N/A**
3	**100 (12/12)**	**88.0 (22/25)**	**100.0 (28/28)**	**100.0 (17/17)**
4	N/A	20.0 (1/5)	37.5 (3/8)	N/A
5	100 (3/3)	25.0 (1/4)	37.5 (3/8)	N/A
Total	86.2 (25/29)	66.1 (37/56)	73.4 (47/64)	96.2 (25/26)
Medium Communities				
6	0 (0/1)	100 (2/2)	100 (2/2)	100 (1/1)
7	**50.0 (1/2)**	**33.3 (2/6)**	**100 (6/6)**	**100 (3/3)**
8	0 (0/1)	66.7 (2/3)	50.0 (1/2)	N/A
9	100 (1/1)	75.0 (3/4)	50.0 (1/2)	100 (1/1)
10	100 (1/1)	100 (2/2)	100 (2/2)	100 (1/1)
Total	50.0 (3/6)	64.7 (11/17)	85.7 (12/14)	100 (6/6)
Far Communities				
11	**100 (3/3)**	**100 (4/4)**	**100 (5/5)**	**100 (3/3)**
12	57.1 (4/7)	50.0 (3/6)	50.0 (1/2)	75.0 (3/4)
13	N/A	100 (1/1)	66.7 (2/3)	100 (1/1)
14	**100 (5/5)**	**73.3 (11/15)**	**88.9 (8/9)**	**83.3 (5/6)**
15	66.7 (2/3)	N/A	N/A	N/A
16	N/A	100 (1/1)	100 (4/4)	60.0 (3/5)
17	70.0 (7/10)	66.7 (8/12)	66.7 (4/6)	100 (12/12)
18	N/A	50.0 (1/2)	100 (2/2)	33.3 (1/3)
19	100 (3/3)	66.7 (2/3)	60.0 (3/5)	33.3 (2/6)
20	N/A	N/A	N/A	25.0 (1/4)
21	N/A	0 (0/2)	33.3 (2/6)	50.0 (4/8)
Total	77.4 (24/31)	67.4 (31/46)	82.1 (32/39)	67.3 (35/52)

	Children receiving one dose (N=47)	Children receiving two doses (N=264)
Records with date information	76.7% (n=36)	73.8% (n=195)
Vaccinated by six months of age	91.7% (n=43)	88.7% (n=173)
Vaccinated by seven months of age	100% (n=46)	96.2% (n=188)

Table C.5: Fraction of children receiving rotarix vaccine by age. These results are shown for all children for whom we had vaccine records, which includes some children who were not in the full regression analysis due to missing data on socioeconomic indicators.

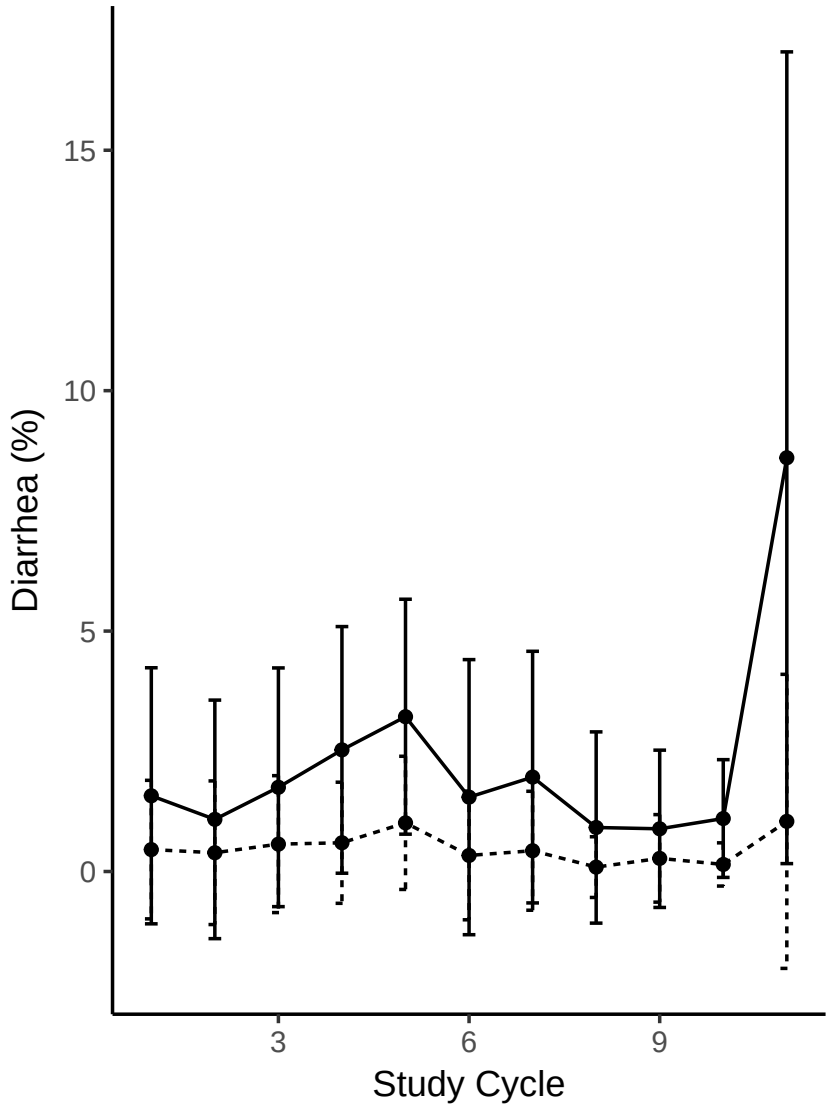

Figure C.2: Prevalence of diarrhea not attributable to rotavirus (solid) and attributable to rotavirus (dashed) over time

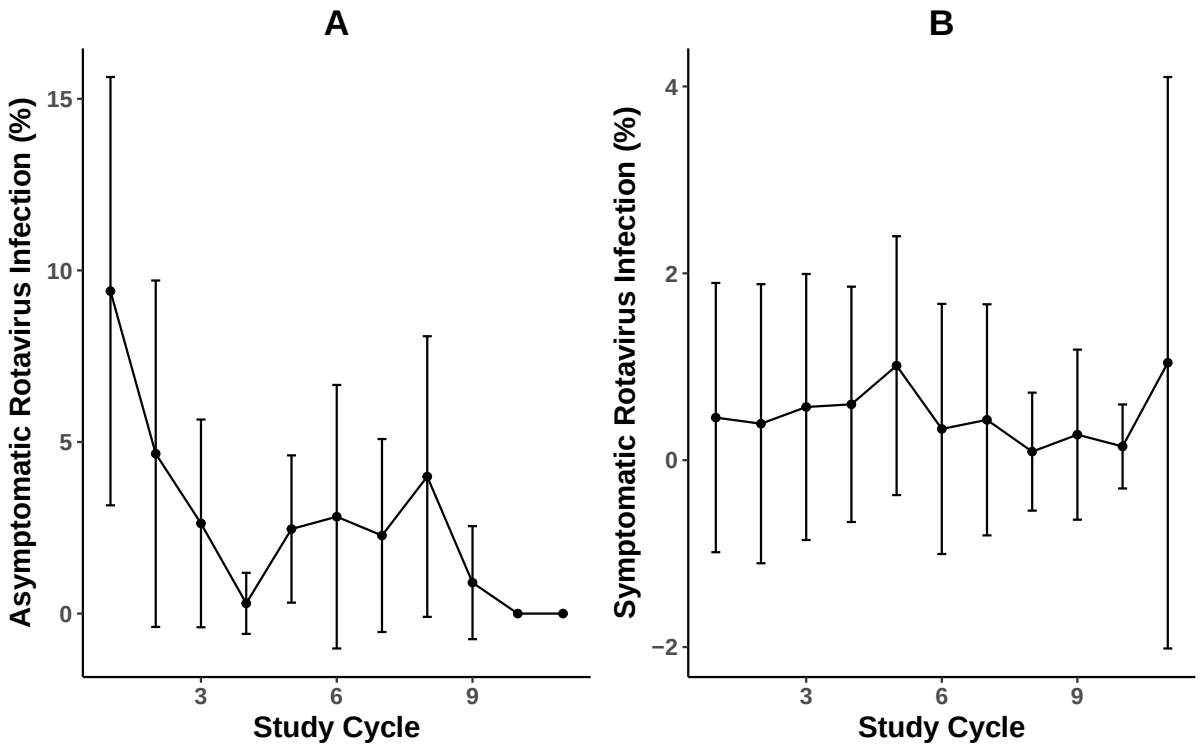

Figure C.3: Prevalence of A) Asymptomatic rotavirus infection and B) Symptomatic rotavirus infection over time. The rotavirus vaccine was introduced in study cycle 8

C.4.2 Additional regression results

C.4.2.1 Socioeconomic status

The socioeconomic indicators used here were not significant in the full population for either rotavirus infection or all-cause diarrhea, but had stronger associations for the youngest age group in the age-stratified model. In general, higher household education and smaller household size would both be expected to indicate higher socioeconomic status.

For the youngest age group, increased household size was associated with increased diarrhea and rotavirus whereas it was protective among older age groups. For this age group, education was a significant predictor of rotavirus infection, with higher education being associated with decreased rotavirus infection. This association was not significant for all-cause diarrhea for any age group. Conceptually, this result is somewhat intuitive because the younger age groups with lower socioeconomic status would tend to get exposed earlier in life and might have increased protection as adults.

C.4.2.2 All-cause diarrhea

The full regression table for the general population is shown in table C.6. After adjusting for relevant confounders, community vaccine coverage was no longer significantly associated with all-cause diarrhea in these six communities.

In this table, model 2 adjusts for rotavirus infection. In the main text (table 4), the vaccine efficacy estimate is not adjusted for whether or not rotavirus was a cause of that diarrhea in order to capture the full effect, given that we are not explicitly modeling synergy between rotavirus and co-infecting pathogens. The vaccine effectiveness estimates for all-cause diarrhea were extremely similar when adjusted for rotavirus infection, but were slightly attenuated and the effect estimates crossed the null.

	Unadjusted	Model 1	Model 2
Rotavirus positive	8.16 (4.92, 13.56)	7.40 (3.29, 14.0)	7.26 (3.83, 13.7)
Male	1.12 (0.914, 1.39)	1.05(0.770, 1.43)	1.05 (0.770, 1.43)
Age			
<1	12.5 (8.27, 18.9)	10.3(6.03, 17.6)	10.3 (6.04, 17.7)
1–5	12.9 (9.87, 16.9)	13.3 (9.73, 18.1)	13.2 (9.72, 18.0)
≥5	Ref	Ref	Ref
Household size	0.967 (0.934, 1.00)	0.965 (0.918, 1.02)	0.965 (0.917, 1.02)
Highest household education	0.949 (0.920, 0.979)	0.971 (0.931, 1.01)	0.972 (0.932, 1.01)
Remoteness			
Close	Ref	Ref	Ref
Medium	1.29 (0.883, 1.87)	1.87 (1.16, 2.99)	1.88 (1.17, 3.02)
Far	0.848 (0.670, 1.07)	1.17 (0.851, 1.60)	1.19 (0.865, 1.62)
Community vaccine coverage	0.672 (0.505, 0.895)	–	0.872 (0.603, 1.26)
Vaccine introduced	0.729 (0.571, 0.931)	–	–

Table C.6: Vaccination coverage and all-cause diarrhea. All cells represent OR (95% CI). Odds ratios are calculated with respect to the reference group for categorical variables. Vaccine coverage is included as a proportion, taking a minimum value of 0 and a maximum value of 1. Model 1 is adjusted for rotavirus infection, gender, age, remoteness, household size, and highest household education. Model 2 is adjusted for all variables in Model 1 plus community vaccine coverage. Note that to calculate vaccine effectiveness, we did not adjust for rotavirus infection and re-ran model 2 without adjusting for this variable.

Age group	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
<1	5.61 (1.24, 25.2)	5.19 (1.16, 23.2)
1-5	5.86 (1.87, 18.3)	5.51 (1.74, 17.5)
≥ 5	10.1 (5.37, 19.0)	10.8 (5.85, 19.8)
Total population	8.20 (4.94, 13.61)	8.69 (4.31, 17.5)

Table C.7: Effect of rotavirus infection on diarrhea illness by age. The adjusted odds ratios are adjusted for sex, household size, highest household education, and remoteness. Each odds ratio can be interpreted as the multiplicative increase in odds of diarrhea, given rotavirus infection.

C.4.2.3 Rotavirus and diarrhea by age group

In order to determine if rotavirus was a causative diarrheal pathogen for all age groups, we also calculated the odds ratio for diarrhea given rotavirus infection relative to the odds of having diarrhea given no rotavirus infection. The results are shown in table C.7. Rotavirus is strongly associated with diarrheal symptoms in all age groups, including older children and adults (age ≥ 5).

C.4.3 Household clustering of rotavirus infection

In order to better understand if infection was occurring within households, we examined how many people who were infected with rotavirus could have been exposed within their households. We did this for all communities, not only communities that were included in the case-control analysis. The resulting diagram is shown in Figure C.4. In our study, we measured a total of 354 rotavirus infections. Of these, we had a stool sample from at least one other household member within the same cycle for 279 (70%). While we did not collect stool samples from all household members, only 25 of these individuals had another household member with a detected infection. While we likely missed some rotavirus infections within these households, this result suggests that, even with sampling error, it is unlikely that the majority of infections could be explained by within-household

transmission. While this result cannot prove that environmental transmission is occurring, it does suggest that a large proportion of rotavirus transmission occurs outside of the household.

It is possible that within-household clustering of infection could be different by age group and/or after the vaccine was introduced. However, we did not have enough secondary cases to investigate this difference. Future studies could collect data from more household members to better evaluate clustering. Ideally, combining household stool samples with environmental surveillance of rotavirus in water and on surfaces could show more definitively the relevance of different transmission routes.

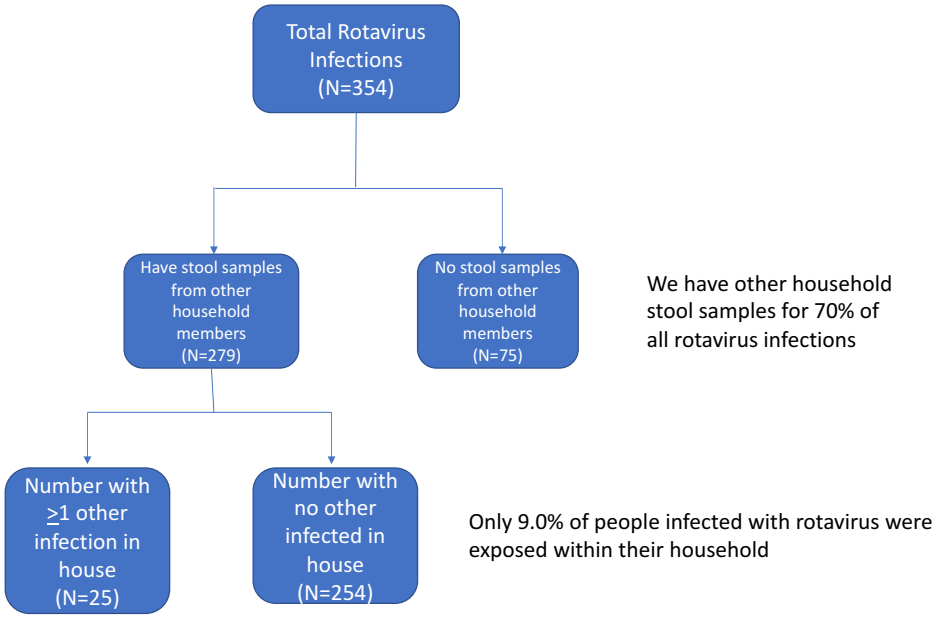

Figure C.4: Schematic diagram illustrating the degree of clustering of rotavirus infections within households

C.5 Supplemental information for attributable fraction analysis (part 2, case only analysis)

A graph of the attributable fraction over time is shown in Figure C.5.

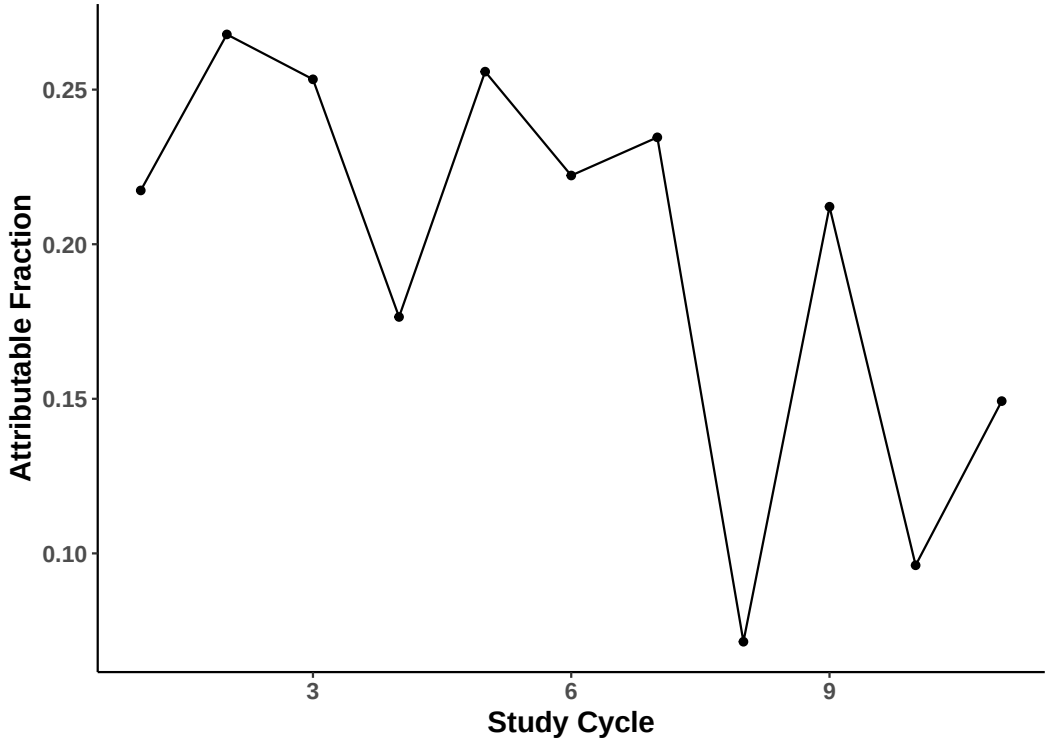

Figure C.5: Fraction of symptomatic diarrhea attributable to rotavirus infection over time (AF)

C.6 Supplemental information for direct effects (part 3, cohort analysis)

The regression results presented in Figure 5 in the main text are also shown in Table C.8 to illustrate differences between the adjusted and unadjusted models. In this population, none of the socioeconomic indicators were significantly associated with all-cause diarrhea, unlike in the case control study. This difference is probably due to the fact that

Table C.8: Poisson Regression Model for the rate of all-cause diarrhea among children. Models 2 and 3 are adjusted for household size, highest household education, gender, community remoteness, and BCG vaccination.

	Model 1 Unadjusted IRR (95% CI)	Model 2 Adjusted IRR (95% CI)	Model 3 Adjusted + Interactions IRR (95% CI)
Rotavirus Vaccine			
0 doses	*Ref*	*Ref*	*Ref*
1 dose	1.22(0.730, 2.04)	0.861 (0.493, 1.50)	0.834 (0.453, 1.53)
2 doses	0.771(0.501, 1.18)	0.619 (0.386, 0.993)	0.491 (0.289, 0.832)
Age $\geq$2 years	0.539(0.374, 0.778)	0.499 (0.344, 0.725)	0.260 (0.087, 0.774)
Male	0.980 (0.721, 1.33)	0.966 (0.704, 1.33)	0.968 (0.705, 1.33)
BCG Vaccination	1.33 (0.654, 2.71)	1.63 (0.747, 3.57)	1.78 (0.808, 3.91)
Household Size	0.98 (0.939, 1.02)	0.976 (0.931, 1.02)	0.971 (0.926, 1.02)
Highest Household Education	1.01 (0.957, 1.06)	0.997 (0.948, 1.05)	1.00 (0.951, 1.05)
Remoteness			
Close	*Ref*	*Ref*	*Ref*
Medium	1.41 (0.852, 2.33)	1.45 (0.870, 2.41)	1.57 (0.942, 2.74)
Far	1.68 (1.21, 2.34)	1.88 (1.32, 2.66)	1.93 (1.35, 2.74)
Age x Vaccine Interaction			
Age($\geq$2) x 1 dose	–	–	1.04 (0.245, 4.39)
Age ($\geq$2) x 2 doses	–	–	2.60 (0.80, 8.46)

children with vaccine records were not comparable to children without vaccine records– children with vaccine records had lower household education than children without vaccine records.

C.7 Internal consistency of results between analysis parts 1 and 3

Prior to doing any analysis, we decided which communities would be included in each part of the analysis in an effort to minimize bias in our vaccine effectiveness estimates. Because we used different communities for different parts of this analysis, we compared the results between them to ensure that our analysis was generalizable to all communities across the region and that the associations we found were not purely the result of our sampling process. In general, our results were highly internally consistent throughout the

	Case-Control Analysis (Part 1) OR (95% CI)	Cohort Analysis (Part 3) IRR (95% CI)
Male	1.04 (0.699, 1.55)	1.05(0.770, 1.43)
Age		
<2	Ref	Ref
2–5	0.406 (0.271, 0.606)	0.519 (0.358, 0.752)
Household size	0.967 (0.898, 1.04)	0.970 (0.926, 1.02)
Highest household education	0.941 (0.885, 1.00)	1.00 (0.951, 1.05)
Remoteness		
Close	Ref	Ref
Medium	1.99 (0.893, 4.44)	1.44 (0.865, 2.39)
Far	0.961 (0.599, 1.54)	1.89 (1.34, 2.66)
Effect of 2 doses*	45.3% (-29.0%, 75.0%)	50.9% (16.8%-71.1%)

*Contrast calculated using the overall coverage parameter for Part 1 and the individual coverage of 2 doses for part 3

Table C.9: Comparison between parts 1 and 3. All parameters except the effect of two doses were calculated using all children in the case-control dataset under 5 years of age for part 1 and all children in the cohort of part 3 and were estimated in a single model. For the effect of two doses, we subset both data files at 2 years of age and presented calculated vaccine effectiveness using the overall coverage parameter for the case control analysis and the two doses for the cohort analysis.

region.

To compare parts 1 and 3, we subset the data from the overall case-control study and compared the results only among children less than five years of age. Because the all-cause diarrhea estimates from the case control study were highly similar both before and after adjusting for vaccine coverage (i.e., the covariates and vaccine coverage were independent predictors of all-cause diarrhea, see Table C.6) we included all study years in this comparison, including time both before and after the vaccine was introduced. By design, the odds ratio approximates the rate ratio in the case control study because controls were time matched to cases and because the outcome was rare in all exposure groups. The results are shown in table C.9.

In general, these results are similar. The main difference is the association between re-

moteness and all-cause diarrhea. In the case control analysis, far communities had similar rates of disease to close communities whereas in the cohort analysis, far communities were at higher risk of disease. This difference most likely arose because children with vaccine records tended to come disproportionately from close/medium villages. Therefore, it is likely that we are not estimating the rate ratio well for far communities in the cohort analysis. The association with education was also slightly different in the cohort analysis, which is also consistent with the marginally significant association we found comparing children with vaccine records to children without vaccine records.

For vaccine coverage, we found somewhat less protection in the case control analysis compared with the cohort. However, for the case-control analysis we excluded all communities with zero rotavirus cases, which limited the effect size we were able to detect. As a result, we would expect the case control analysis to estimate somewhat less protection than the cohort analysis. Despite this difference, the results are quite similar, with less than a 5% difference in estimated vaccine effectiveness.

While not directly comparable to any of our other analyses, it is encouraging that the results from part 2 (attributable fraction analysis) also demonstrated strong associations with vaccination status that were similar to the effects on symptomatic rotavirus infection in part 1 (53.2% efficacy against symptomatic rotavirus compared with a 54.5% decrease in the attributable fraction). The results from part 2 are not exactly parallel to the case control analysis because in part 2 we assess the impact of rotavirus vaccination among cases whereas in part 1 we assess the impact of vaccination among both cases and controls (i.e., the denominator for the calculation is different). Because rotavirus infection is common among cases, we would generally expect any effects among cases of diarrhea to be larger than those among the full population, which is exactly what we see.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Levy K, A.P W, Goldstein R, Carlton E (2016) Untangling the impacts of climate change on waterborne diseases: a systematic review of relationships between diarrheal diseases and temperature, rainfall, flooding, and drought. *Environmental Science and Technology* 50:4905–4922.

[2] Panozzo CA et al. (2014) Direct, indirect, total, and overall effectiveness of the rotavirus vaccines for the prevention of gastroenteritis hospitalizations in privately insured US children, 2007–2010. *American journal of epidemiology* 179(7):895–909.

[3] de Blasio BF, Kasymbekova K, Flem E (2010) Dynamic model of rotavirus transmission and the impact of rotavirus vaccination in kyrgyzstan. *Vaccine* 28(50):7923–32.

[4] Kotloff K et al. (2013) 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. *The Lancet* 382(9888):209–222.

[5] Abubakar I, Tillmann T, Banerjee A (2015) Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. *Lancet* 385(9963):117–171.

[6] Murray CJ et al. (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The Lancet* 380(9859):2197–2223.

[7] Walker CLF et al. (2013) Global burden of childhood pneumonia and diarrhoea. *The Lancet* 381(9875):1405–1416.

[8] Lozano R et al. (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *The Lancet* 380(9859):2095–2128.

[9] Tate JE et al. (2012) 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. *The Lancet infectious diseases* 12(2):136–141.

[10] Platts-Mills JA et al. (2015) Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (mal-ed). *The Lancet Global health* 3(9):e564–e575.

[11] Greenberg HB, Estes MK (2009) Rotaviruses: from pathogenesis to vaccination. *Gastroenterology* 136(6):1939–1951.

[12] Santos N, Hoshino Y (2005) Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Reviews in medical virology* 15(1):29–56.

[13] Gentsch JR et al. (2005) Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. *Journal of Infectious Diseases* 192(Supplement_1):S146–S159.

[14] Gurgel RQ et al. (2007) Predominance of rotavirus P [4] G2 in a vaccinated population, Brazil. *Emerging infectious diseases* 13(10):1571.

[15] Zeller M et al. (2010) Rotavirus incidence and genotype distribution before and after national rotavirus vaccine introduction in Belgium. *Vaccine* 28(47):7507–7513.

[16] Patel MM et al. (2011) Real-world impact of rotavirus vaccination. *The Pediatric infectious disease journal* 30(1):S1–S5.

[17] Chiba S et al. (1986) Protective effect of naturally acquired homotypic and heterotypic rotavirus antibodies. *The Lancet* 328(8504):417–421.

[18] Velázquez FR et al. (1996) Rotavirus infection in infants as protection against subsequent infections. *New England Journal of Medicine* 335(14):1022–1028.

[19] Anderson E, S.G. W (2004) Rotavirus infection in adults. *The Lancet Infectious Diseases* 4:91–99.

[20] Bosch A (1998) Human enteric viruses in the water environment: a minireview. *Int Microbiol* 1(3):191–6.

[21] Gerba CP (2000) Assessment of enteric pathogen shedding by bathers during recreational activity and its impact on water quality. *Quantitative Microbiology* 2(1):55–68.

[22] Mukhopadhya I et al. (2013) Rotavirus shedding in symptomatic and asymptomatic children using reverse transcription-quantitative pcr. *Journal of medical virology* 85(9):1661–1668.

[23] Moe K, Shirley JA (1982) The effects of relative humidity and temperature on the survival of human rotavirus in faeces. *Archives of Virology* 72:179–186.

[24] Sattar S, N LE, Springthorpe V, Nair R (1986) Institutional outbreaks of rotavirus diarrhoea: potential role of fomites and environmental surfaces as vehicles for virus transmission. *J Hyg* 96:277–289.

[25] Ansari S, Springthorpe V, Sattar S (1991) Survival and vehicular spread of human rotaviruses: possible relation to seasonality of outbreaks. *Rev Infec Dis* 13(3):448–461.

[26] Boone S, Gerba C (2007) Significance of fomites in the spread of respiratory and enteric viral disease. *Applied and Environmental Microbiology* 73(6):1687–1696.

[27] Abad FX, Pintó RM, Bosch A (1994) Survival of enteric viruses on environmental fomites. *Applied and Environmental Microbiology* 60:3704–3710.

[28] Espinosa A et al. (2008) Infectivity and genome persistence of rotavirus and astrovirusin groundwater and surface water. *Water Research* 42(10-11):2618–2628.

[29] Ward R, Knowlton D, Winston P (1986) Mechanism of inactication of enteric viruses in fresh water. *Applied and Environmental Microbiology* 52(3):450–459.

[30] Raphael R, Sattar S, Springthorpe V (1985) Long-term survival of human rotavirus in raw and treated river water. *Canadian Journal of Microbiology* 31(2):124–128.

[31] El-Sanousy W, Osman G, Melegy A (2014) Survival of adnovirus, rotavirus, hepatitus A virus, pathogenic bacteria, and bacterial indicators in ground water. *World Applied Sciences Journal* 29(3):337–348.

[32] Hansen J, Warden P, Margolin A (2007) Inactication of adenovirus type 5, rotavirus WA and male specific coliphage (MS2) in biosolids by lime stabilization. *International Journal of Environmental Research and Public Health* 4(1):61–67.

[33] Sattar S, Raphael R, Springthorpe V (1984) Rotavirus survival in conventionally treated drinking water. *Canadian Journal of Microbiology* 30(5):653–656.

[34] Höglund C, Ashbolt N, Stenström T, Svenson L (2002) Viral persistence in source-separated urine. *Advances in Environmental Research* 6:265–275.

[35] Chung H, Sobsey M (1993) Comparative survival of indicator viruses and enteric viruses in seawater and sediment. *Water Science and Technology* 27(3-4):425–428.

[36] McDaniels A, Cochran K, Gannon J, Williams G (1983) Rotavirus and reovirus stability in microorganism-free distilled and wastewaters. *Water Research* 17(10):1349–1353.

[37] Bhavnani D, Goldstick JE, Cevallos W, Trueba G, Eisenberg JN (2012) Synergistic effects between rotavirus and coinfecting pathogens on diarrheal disease: evidence from a community-based study in northwestern ecuador. *American journal of epidemiology* 176(5):387–395.

[38] Ward R et al. (1986) Human rotavirus studies in volunteers: determination of infectious dose and serological response to infection. *Journal of Infectious Diseases* 154(5):871–880.

[39] Pitzer V et al. (2009) Demographic variability, vaccination, and the spatiotemporal dynamics of rotavirus epidemics. *Science* 325(5938):290–294.

[40] Pitzer V et al. (2011) Influence of birth rates and transmission rates on the global seasonality of rotavirus incidence. *Journal Royal Society Interface* 8:1584–1593.

[41] Martinez P, King A, Yunus M, Faruque A, Pascual M (2016) Differential and enhanced response to cliamte forcing in diarrheal disease due to rotavirus across a megacity of the developing world. *Proceedings of the National Academy of Sciences* 113(15):4092–4097.

[42] Hrdy DB (1987) Epidemiology of rotaviral infection in adults. *Reviews of infectious diseases* 9(3):461–469.

[43] Levy K, Hubbard A, Eisenberg J (2009) Seasonality of rotavirus in the tropics: a systematic review and meta-analysis. *International Journal of Epidemiology* 38(6):1487–1496.

[44] Carlton E, Woster A, DeWitt P, Goldstein R, Levy K (2016) A systematic review and meta-analysis of ambient temperature and diarrhoeal diseases. *International Journal of Epidemiology* 45(1):117–130.

[45] Cook SM, Glass RI, LeBaron CW, Ho M (1990) Global seasonality of rotavirus infections. *Bulletin of the World Health Organization* 68(2):1281–1289.

[46] Ruan S, Wang W, Levin SA (2005) The effect of global travel on the spread of sars. *Mathematical Biosciences and Engineering* 3(1):205–218.

[47] Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ (2006) Delaying the international spread of pandemic influenza. *PLoS Medicine* 3:e212.

[48] Grais RF, Ellis JH, Glass GE (2003) Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology* 18(11):1065–1072.

[49] Lee K, Dodgson R (2000) Globalization and cholera: implications for global governance. *Global Governance* 6(2):213–236.

[50] Wesolowski A et al. (2014) Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phone. *Scientific reports* 4.

[51] Riley S (2007) Large-scale spatial-transmission models of infectious disease. *Science (New York, N.Y.)* 316(5829):1298–1301.

[52] Simini F, González MC, Maritan A, Barabasi AL (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96–100.

[53] González MC, Hidalgo CA, Albert-László B (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779–782.

[54] O'Ryan M, Lucero Y, Linhares AC (2011) Rotarix®: vaccine performance 6 years postlicensure. *Expert review of vaccines* 10(12):1645–1659.

[55] Vesikari T et al. (2007) Efficacy of human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in european infants: randomised, double-blind controlled study. *The Lancet* 370(9601):1757–1763.

[56] Phua KB et al. (2009) Safety and efficacy of human rotavirus vaccine during the first 2 years of life in asian infants: randomised, double-blind, controlled study. *Vaccine* 27(43):5936–5941.

[57] Linhares AC et al. (2008) Efficacy and safety of an oral live attenuated human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in latin american infants: a randomised, double-blind, placebo-controlled phase iii study. *The Lancet* 371(9619):1181–1189.

[58] Madhi SA et al. (2010) Effect of human rotavirus vaccine on severe diarrhea in african infants. *New England Journal of Medicine* 362(4):289–298.

[59] Naylor C et al. (2015) Environmental enteropathy, oral vaccine failure and growth faltering in infants in Bangladesh. *EBioMedicine* 2(11):1759–1766.

[60] Serazin AC, Shackelton LA, Wilson C, Bhan MK (2010) Improving the performance of enteric vaccines in the developing world.

[61] Cortese M, Parashar U (2009) Prevention of rotavirus gastroenteritis among infants and children recommendation of the advisory committee on immunization practices. *MMWR* 58(RR02):1–25.

[62] World Health Organization (2014) Immunization summary: a statistical reference containing data through 2013, Technical report.

[63] Madsen LB, Ustrup M, Fischer TK, Bygbjerg IC, Konradsen F (2012) Reduced price on rotavirus vaccines: enough to facilitate access where most needed? *Bulletin of the World Health Organization* 90(7):554–556.

[64] Parez N et al. (2014) Rotavirus vaccination in Europe: drivers and barriers. *The Lancet Infectious Diseases* 14(5):416–425.

[65] Tate JE et al. (2011) Uptake, impact, and effectiveness of rotavirus vaccination in the United States: review of the first 3 years of postlicensure data. *The Pediatric Infectious Disease Journal* 30(1):S56–S60.

[66] Yen C et al. (2011) Decline in rotavirus hospitalizations and health care visits for childhood diarrhea following vaccination in el salvador. *Pediatric infectious disease journal* 30(1):S6–S10.

[67] Payne DC et al. (2011) Direct and indirect effects of rotavirus vaccination upon childhood hospitalizations in 3 US counties, 2006-2009. *Clinical Infectious Diseases* 53(3):245–253.

[68] Eisenberg JNS et al. (2006) Environmental change and infectious disease: how new roads affect the transmission of diarrheal pathogens in rural ecuador. *Proceedings of the National Academy of Sciences of the United States of America* 103(51):19460–5.

[69] Rao G et al. (2015) Spatial variability of Escherichia coli in rivers of northern coastal Ecuador. *Water* 7(2):818–832.

[70] Prothero RM (1977) Disease and mobility: a neglected factor in epidemiology. *International journal of epidemiology* 6(3):259–267.

[71] Tatem AJ, Rogers DJ, Hay SI (2006) Global transport networks and infectious disease spread. *Advances in parasitology* 62:293–343.

[72] Cliff A (2004) Time, travel and infection. *British Medical Bulletin* 69(1):87–99.

[73] Martens P, Hall L (2000) Malaria on the move : human population movement and malaria transmission. *Emerging Infectious Diseases* 103(April):1–7.

[74] Stoddard ST et al. (2009) The role of human movement in the transmission of vector-borne pathogens. *PLoS Neglected Tropical Diseases* 3(7).

[75] Tuite AR et al. (2011) Cholera epidemic in Haiti, 2010: Using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Annals of Internal Medicine* 154(9):593–601.

[76] Viboud C et al. (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science (New York, N.Y.)* 312(5772):447–451.

[77] Stoddard ST et al. (2013) House-to-house human movement drives dengue virus transmission. *Proceedings of the National Academy of Sciences of the United States of America* 110(3):994–9.

[78] Windle J, Cramb RA (1997) Remoteness and rural development: economic impacts of rural roads on upland farmers in Sarawak, Malaysia. *Asia Pacific Viewpoint* 38(1):37–53.

[79] Rudel TK, Richards S (1990) Urbanization, roads, and rural population change in the Ecuadorian Andes. *Studies in Comparative International Development* 25(3):73–89.

[80] Zelner JL et al. (2012) Social connectedness and disease transmission: social organization, cohesion, village context, and infection risk in rural ecuador. *American journal of public health* 102(12):2233–9.

[81] Vittor AY et al. (2006) The effect of deforestation on the human-biting rate of Anopheles darlingi, the primary vector of falciparum malaria in the peruvian amazon. *The American journal of tropical medicine and hygiene* 74(1):3–11.

[82] Dutta P et al. (1998) Distribution of potential dengue vectors in major townships along the national highways and trunk roads of northest india. *Southeast Asian Journal of Tropical Medicine Public Health* 29:173–176.

[83] Crépey P, Barthélemy M (2007) Detecting robust patterns in the spread of epidemics: A case study of influenza in the united states and france. *American Journal of Epidemiology* 166(11):1244–1251.

[84] Balcan D et al. (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America* 106(51):21484–9.

[85] Gatto M et al. (2013) Spatially explicit conditions for waterborne pathogen invasion. *The American Naturalist* 182(3):328–346.

[86] Gatto M et al. (2012) Generalized reproduction numbers and the prediction of patterns in waterborne disease. *Proceedings of the national academy of sciences* 109(48):19703–19708.

[87] Knipl D (2016) A new approach for designing disease intervention strategies in metapopulation models. *J Biol Dyn* 10:71–94.

[88] Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. *Nature* 438(7066):355–9.

[89] Stein RA (2011) Super-spreaders in infectious diseases. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 15(8):e510–3.

[90] Widdowson MA, Steele D, Vojdani J, Wecker J, Parashar U (2009) Global rotavirus surveillance: determining the need and measuring the impact of rotavirus vaccines. *J Infect Dis* 200 Suppl 1:S1–8.

[91] Dietz K (1993) The estimation of the basic reproductive number for infectious diseases. *Statistical Methods in Medical Research* 2:23–41.

[92] Hanley JA, Negassa A, Forrester JE, et al. (2003) Statistical analysis of correlated data using generalized estimating equations: an orientation. *American journal of epidemiology* 157(4):364–375.

[93] Pan W (2001) Akaike's information criterion in generalized estimating equations. *Biometrics* 57(1):120–125.

[94] Gushulak BD, MacPherson DW (2004) Globalization of infectious diseases: the impact of migration. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 38(12):1742–8.

[95] Glass RI et al. (2006) Rotavirus vaccines: current prospects and future challenges. *The Lancet* 368(9532):323–332.

[96] Bishop RF, Barnes GL, Cipriani E, Lund JS (1983) Clinical immunity after neonatal rotavirus infection. *New England Journal of Medicine* 309(2):72–76. PMID: 6304516.

[97] Montgomery MR (2008) The urban transformation of the developing world. *Science* 319(5864):361–364.

[98] Kassebaum NJ, Arora M, Barber R, et al. (2016) Global, regional, and national disability adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the global burden of disease study 2015. *Lancet* 388:1603–1658.

[99] Dennehy P (2013) Transmission of rotavirus and other infectious pathogens in the home. *Pediatric Infectious Disease Journal* 19:S103–105.

[100] Kolstad E, Johansson K (2011) Uncertainties associated with quantifying climate change impacts on human health: a case study for diarrhea. *Environmental Health Perspectives* 119(3):299–305.

[101] Sharpiro RS, Cowen LE (2012) Thermal control of microbial development and virulence: mechanisms of microbial temperature sensing. *mBio* 3(5):e00238–12.

[102] Konkel M, Tilly K (2000) Temperature regulated expression of bacterial virulence genes. *Microbes and infection* 2(2):157–66.

[103] Mellor J et al. (2016) Planning for climate change: the need for mechanistic systems-based approaches to study climate change and impacts on diarrheal diseases. *Science of the Total Environment* 548-549:82–90.

[104] Trummers B (2006) Datathief iii. *http://datathief.org/*.

[105] Crane S, Moore J (1986) Modeling enteric bacterial die off: a review. *Water, air, and soil pollution* 27:411–439.

[106] Brouwer AF et al. (2017) Modeling biphasic environmental decay of pathogens and implications for risk analysis. *Environmental Science and Technology* 51:2186–2196.

[107] Peng X, Murphy T, Holden N (2008) Evaluation of the effect of temperature on the die-off rate for cryptosporidium parvum oocysts in water, soils, and feces. *Appl Environ Microbiol* 74(23):7101–7107.

[108] Bertuzzo E et al. (2008) On the space-time evolution of a cholera epidemic. *Water Resources Research* 44:W01424.

[109] Downing J et al. (2006) The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol Oceanogr* 51(5):2388–2397.

[110] public health C, environmental engineering organization (1999) Manual of water supply and treatment.

[111] Teunis P, van der Heijden O, van der Giessen J, Havelaar A (1996) The dose-response relation in human volunteers for gastro-intestinal pathogens. *Rijksinstituut voor volksgezondhid en milieu bilthoven* 284550002:1–87.

[112] Brouwer AF, Weir MH, Eisenberg MC, Reza R, Eisenberg JNS (2017) Dose-response relationships for environmentally mediated infectious disease transmission models. *PLoS Computational Biology* 13:e1005481.

[113] Banks R (1974) A mixing cell model for longitudinal dispersion in open channels. *Water Resources Research* 10:357–358.

[114] U.S. Environmental Protection Agency (2011) Exposure factors handbook 2011 edition (final), (Washington, DC, EPA/600/R-09/052F), Technical report.

[115] Gurwith M, Wenman W, Hinde D, Feltham S, Greenberg H (1981) A prospective study of rotavirus infection in infants and young children. *Journal of Infectious Diseases* 144(3):218–224.

[116] Molla A, Rahman M, Sarker S, Sack D, Molla A (1981) Stool electrolyte content and purging rates in diarrhea caused by rotavirus, enterotoxigenic e. coli, and v. cholerae in children. *Journal of Pediatrics* 98(5):835–838.

[117] Fang S et al. (2009) Dose-dependent effect of lacobacillus rhamnosus on quantitative reduction of faecal rotavirus shedding in children. *Journal of Tropical Pediatrics* 55(5):297–301.

[118] WHO (2006) World health organization child growth standards: length/height for age, weight-for-age, weight-for-length, weight-for-height, and body mass index-for-age. pp. 1–312.

[119] Van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical biosciences* 180(1):29–48.

[120] Tien J, Earn D (2010) Multiple transmission pathways and disease dynamics in a waterborne pathogen model. *Bulletin of Mathematical Biology* 72:1506–1533.

[121] Nishiura H, Chowell G (2009) The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. *Mathematical and statistical estimation approaches in epidemiology* pp. 103–121.

[122] Patel M et al. (2013) Global seasonality of rotavirus disease. *Pediatr Infect Dis J* 32(4):e134–e147.

[123] Rothman K, Greenland S (2005) Causation and causal inference in epidemiology. *American Journal of Public Health* 95:S144–S150.

[124] Andreason V (2011) The final size of an epidemic and its relation to the basic reproduction number. *Bulletin of Mathematical Biology* 73:2305–2321.

[125] Eisenberg J, Lewis B, Porco T, Hubbard A, Colford J (2003) Bias due to secondary transmission in estimation of attributable risk from intervention trials. *Epidemiology* 14(4):442–450.

[126] Glass R et al. (2012) The control of rotavirus gastroenteritis in the United States. *Transactions of the American Clinical and Climatological Association* 123:36–53.

[127] Intergovernmental Panel on Climate Change (2001) Climate change 2001: The scientific basis. *Contribution of Working Group I to the Third Assessment Report* pp. 1–944.

[128] Deatherage DE, Kepner JL, Bennett AF, Lenski RE, Barrick JE (2017) Specificity of genome evolution in experimental populations of Escherichia coli evolved at different temperatures. *Procedings of the National Academy of Sciences* Published online February 15, 2017:E1904–E1912.

[129] Dennehy PH (2008) Rotavirus vaccines: an overview. *Clinical microbiology reviews* 21(1):198–208.

[130] O'Ryan M, Linhares AC (2009) Update on Rotarix: an oral human rotavirus vaccine. *Expert review of vaccines* 8(12):1627–1641.

[131] Jonesteller CL, Burnett E, Yen C, Tate JE, Parashar UD (2017) Effectiveness of rotavirus vaccination: A systematic review of the first decade of global postlicensure data, 2006–2016. *Clinical Infectious Diseases* 65(5):840–850.

[132] Ustrup M, Madsen LB, Bygbjerg IC, Konradsen F (2011) Outstanding challenges for rotavirus vaccine introduction in low-income countries–a systematic review. *Danish Med Bull* 58(10):A4323.

[133] Justino MCA et al. (2011) Effectiveness of the monovalent G1P [8] human rotavirus vaccine against hospitalization for severe G2P [4] rotavirus gastroenteritis in belem, brazil. *The Pediatric infectious disease journal* 30(5):396–401.

[134] Correia JB et al. (2010) Effectiveness of monovalent rotavirus vaccine (Rotarix) against severe diarrhea caused by serotypically unrelated G2P [4] strains in Brazil. *The Journal of infectious diseases* 201(3):363–369.

[135] de Palma O et al. (2010) Effectiveness of rotavirus vaccination against childhood diarrhoea in El Salvador: case-control study. *Bmj* 340:c2825.

[136] Franco MA, Angel J, Greenberg HB (2006) Immunity and correlates of protection for rotavirus vaccines. *Vaccine* 24(15):2718–2731.

[137] Anderson EJ (2008) Rotavirus vaccines: viral shedding and risk of transmission. *The Lancet Infectious Diseases* 8(10):642–649.

[138] Kraay ANM, Trostle J, Brouwer AF, Trujillo WC, Eisenberg JNS (2017) Determinants of short-term movement in a developing region and implications for disease transmission. *Epidemiology*.

[139] Aaby P, Kollmann TR, Benn CS (2014) Nonspecific effects of neonatal and infant vaccination: public-health, immunological and conceptual challenges. *Nature immunology* 15(10):895–899.

[140] Halloran ME, Hudgens MG (2017) Estimating population effects of vaccination using large, routinely collected data. *Statistics in Medicine*.

[141] Ruiz-Palacios GM et al. (2006) Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. *New England Journal of Medicine* 354(1):11–22.

[142] Prendergast AJ, Humphrey JH (2014) The stunting syndrome in developing countries. *Paediatrics and International Child Health* 34(4):250–265.

[143] Wesolowski A et al. (2014) Commentary: containing the ebola outbreak-the potential and challenge of mobile network data. *PLoS currents* 6.

[144] Guyant P et al. (2015) Malaria and the mobile and migrant population in Cambodia: a population movement framework to inform strategies for malaria control and elimination. *Malaria journal* 14(1):252.

[145] Anderson EJ, Weber SG (2004) Rotavirus infection in adults. *The Lancet infectious diseases* 4(2):91–99.

[146] Keusch GT et al. (2014) Environmental enteric dysfunction: pathogenesis, diagnosis, and clinical consequences. *Clinical Infectious Diseases* 59(suppl_4):S207–S212.

[147] Carvalho-Costa FA et al. (2011) Laboratory-based rotavirus surveillance during the introduction of a vaccination program, Brazil, 2005–2009. *The Pediatric infectious disease journal* 30(1):S35–S41.

[148] Nakagomi T et al. (2008) Apparent extinction of non-G2 rotavirus strains from circulation in Recife, Brazil, after the introduction of rotavirus vaccine. *Archives of virology* 153(3):591–593.

[149] Hasing ME et al. (2009) Rapid changes in rotaviral genotypes in Ecuador. *Journal of medical virology* 81(12):2109–2113.

[150] Solberg OD, Hasing ME, Trueba G, Eisenberg JNS (2009) Characterization of novel vp7, vp4, and vp6 genotypes of a previously untypeable group a rotavirus. *Virology* 385(1):58–67.

[151] Maudlin I, Eisler MC, Welburn SC (2009) Neglected and endemic zoonoses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364(1530):2777–2787.

[152] Moss WJ, Strebel P (2011) Biological feasibility of measles eradication. *The Journal of infectious diseases* 204(suppl_1):S47–S53.

[153] Bharti N et al. (2010) Measles hotspots and epidemiological connectivity. *Epidemiology & Infection* 138(9):1308–1316.