Spring 5-10-2018

Characterizing Multiple Spatial Waves of the 1991-1997 Cholera Epidemic in Peru

Natalie Sterrett

ABSTRACT

CHARACTERIZING MULTIPLE SPATIAL WAVES OF THE 1991-1997 CHOLERA EPIDEMIC IN PERU

By

NATALIE STERRETT

4/23/17

Background

Due to a lack of sanitary infrastructure and a highly susceptible population, Peru experienced a historic outbreak of Vibrio cholerae O1 that began in 1991 and generated multiple waves of disease for several years. Though case-fatality was low, the epidemic put massive strain on healthcare and governmental resources. Here we explore the transmission dynamics and spatiotemporal variation of cholera in Peru using mathematical models and statistical analyses that account for environmental conditions favoring the persistence of bacteria in the environment.

Methods

The authors use dynamic transmission models that incorporate seasonal variation in temperature, concentration of vibrios in the environment, as well as separate human and environmental transmission pathways. The model is fit to weekly department level data obtained from the cholera surveillance system in Peru. The authors also assess the spatial patterns of cholera transmission and correlations between case incidence, time of epidemic onset, and department level variables. Reproductive numbers are compared across departments.

Results

Our findings indicate that the epidemic first hit the coastal departments of Peru and later spread through the highlands and jungle regions. There was high seasonal variation in case incidence, with three clear waves of transmission corresponding to the warm seasons in Peru. Department level variables such as population size and elevation also played a role in transmission patterns. Finally, basic reproductive numbers most often ranged from one to eleven depending on department and time of year. Lima had the largest reproductive number, likely due to its population density and proximity to the coast.

Conclusions

Incorporating environmental variables into an epidemic model predicts the multiple waves of transmission characteristic of *V. cholerae*, and effectively differentiates transmission patterns by

geographic region even in the absence of unique parameter estimates. Mathematical models can provide valuable information about transmission patterns and should continue to be used to inform public health decision making.

CHARACTERIZING MULTIPLE SPATIAL WAVES OF THE 1991-1997 CHOLERA EPIDEMIC IN PERU

by

NATALIE STERRETT

B.S., EMORY UNIVERSITY

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA 30303

APPROVAL PAGE

CHARACTERIZING MULTIPLE SPATIAL WAVES OF THE 1991-1997 CHOLERA EPIDEMIC IN PERU

by

NATALIE STERRETT

Approved:

Gerardo Chowell, PhD
Committee Chair

Kimberlyn Roosa, MPH
Committee Member

April 23, 2018
Date

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Natalie Sterrett
Signature of Author

TABLE OF CONTENTS

List of Tables

List of Figures

1 Background

Vibrio cholerae continues to generate outbreaks of acute gastrointestinal illness particularly in lower-income countries with poor sanitary infrastructure; it currently affects 1.3 to 4 million people annually worldwide [1]. Transmitted through contaminated food or water, cholera poses a major public health threat in many countries around the world [2]. Ecological studies have confirmed warm brackish waters to be an ideal reservoir for *V. cholerae*, where it can attach to aquatic organisms such as shellfish and zooplankton [3, 4, 5]. This bacterium thrives at warmer temperatures which increases the risk of cholera outbreaks among susceptible populations [2]. Although our understanding of this pathogen has advanced considerably, there are still open questions relating to cholera persistence and transmission dynamics at different spatial scales. To improve prevention and mitigation strategies, we must continue to develop our understanding of the transmission dynamics of cholera through the lens of past epidemics.

Historically, cholera epidemics have been associated with one of two biotypes, both of which belong to *V. cholerae* serogroup O1 [2]. Up to 1960, epidemic cholera was primarily caused by the classical O1 biotype, but it was subsequently replaced by *V. cholerae* El Tor, which marked the beginning of the seventh cholera pandemic in 1961 [2, 6, 7]. By the early 1990s, an effective vaccine had yet to become widely available [8, 9, 10, 11]. Fortunately, infection with the El Tor biotype most often results in mild or asymptomatic infection, an evolutionary tradeoff for its increased ability to survive in both human hosts and the environment [12]. However, in approximately one out of every 10-50 infected individuals it is characterized by the rapid onset of watery diarrhea, vomiting, cramping, and subsequent dehydration [2, 13, 14]. If left untreated, it can lead to shock, renal failure, and eventually death [2, 15]. Proper intravenous and oral rehydration, along with appropriate use of antibiotics, can reduce case-fatality to less than 1% [15].

The ongoing seventh cholera pandemic, caused by an El Tor strain originating in the Bay of Bengal, involves at least three separate waves of transmission [16]. The pandemic got its start in Indonesia in 1961 [17], from there spreading through India (1964) [18], Africa (1970) [19, 20], southern Europe (1970) [21, 22], and South America (1991) [23, 24, 25]. Thereafter, Peru experienced one of the worst multi-year epidemics in South American history for which an effective vaccine was unavailable. The first cases were reported in Peru during the final week of January of 1991, and these cases marked the onset of a brutal assault on the South American population. This epidemic likely caused close to a million cases and almost 9,000 deaths between January 1991 and December 1993 [26, 27]. Although it was initially speculated that a Chinese ship imported the disease through an infected crew member [28], it is plausible that the culprit pathogen was already widespread in the local environment, leading to sporadic cholera cases as early as October 1990 and implicating an environmental source for the epidemic [29].

A limited picture of the impact of the epidemic in Peru can be gleaned through descriptive local reports of the epidemic [11, 25, 30, 31, 32, 33]. For instance, some reports suggest that the first cases of cholera in Peru were reported in the central coast in late January of 1991 [34], whereas subsequent cases were reported almost simultaneously in coastal cities farther north [33]. The disease was then reported rapidly in the rest of the country [24]. Cholera continued to generate outbreaks

for several more years [35]. Nevertheless, a quantitative analysis of the spatiotemporal spread of this devastating cholera epidemic in Peru is sorely needed in order to understand the transmission dynamics of cholera in a mostly susceptible population and in the absence of an effective vaccine.

Mathematical modeling studies of cholera epidemics have been useful to quantify transmission rates and reproductive numbers for epidemics in various areas [36, 37, 38, 39]. However, we argue that the 1991-1997 cholera epidemic in Peru is an interesting case study as it involved multiple waves of disease that affected a spatially heterogeneous population at a time when a cholera vaccine was unavailable. In this study we examine the geographic distribution and temporal variation of this major cholera epidemic using mathematical modeling and statistical analyses together with a unique dataset of time series of cholera cases across departments in Peru. We seek to characterize transmission patterns across geographic regions, estimate reproductive numbers for both environment to human and human to human transmission pathways, and to evaluate the effects of environmental variables and temporal trends in cholera incidence at different spatial scales: national, regional, and department. Finally, an appropriate model is applied to examine transmission dynamics and evaluate the relative importance of environmental and human transmission routes for *V. cholerae* in Peru.

2 Methods

2.1 Setting

Peru is located on the Pacific coast, sharing borders with Bolivia, Brazil, Chile, Colombia, and Ecuador. In 1990, the total population in Peru exceeded 22 million, heterogeneously distributed in a surface area of 1,285,220 km^2. The geographic setting of Peru is unique compared to other cholera affected countries. The country is divided into 25 administrative regions and is comprised of three geographic zones with varying climates: the western coast along the Pacific Ocean, the highlands or more central departments, and the Amazon jungle toward the East. Respectively, the climates are dry, temperate, and tropical. In a country like Peru, it is important to differentiate between the geographic regions, as cholera transmission rates from the environmental route are temperature dependent [3].

While cholera rarely presents itself in high-income countries, the lack of economic resources and basic sanitation in Peru allowed cholera to wreak havoc on a susceptible population in 1991 [25, 28]. At the time, Peru struggled with access to healthcare, environmental sanitation, and the resources necessary to mitigate a large epidemic [35]. In addition, inadequate treatment of water throughout the country, as well as deficiencies in water storage systems, have been documented [11, 40]. Further, the political and economic climate in the early 1990s hampered mitigation efforts [11]. Given the great magnitude of the epidemic, the healthcare infrastructure was overrun by the high number of cases presenting to clinics and hospitals [41].

2.2 Data

2.2.1 Epidemiologic Data

Peru's General Office of Epidemiology launched the epidemiological surveillance system with weekly reporting in 1989, not long before the cholera epidemic hit Peru [42]. This surveillance system generates weekly surveillance reports across 25 departments by relying on a network of over 6,000 geographically distributed notifying units [43].

During the 1991-1997 cholera epidemic, surveillance included both confirmed and suspected cases. Confirmed cases were laboratory confirmed with *V. cholerae* 01 El Tor Inaba [44, 45]. The definition of suspected cases was stable during the epidemic and consisted of cases defined as individuals greater than five years of age presenting with acute and watery diarrhea after the epidemic onset [44, 46].

2.2.2 Environmental and Geographic Data

Weekly temperature time series from 1991 to 1997 were obtained from the European Centre for Medium Range Weather Forecasts ERA-Interim atmospheric reanalysis archive. The ERA-interim model allows estimation of daily minimum, mean, and maximum temperatures by department [47]. Model estimates were used to assess correlations between case incidence and temperature at the department level. Weekly mean, minimum, and maximum temperatures by department can be

seen in Figure 1. We also collected several geographic variables at the department level including elevation, population size, latitude, and longitude (Table S1), which were used in spatial analyses.

2.3 Spatial Analysis

To evaluate elevation as a proxy for temperature, we used Spearman's rho (ρ) to assess correlations between cumulative incidence and department elevation. We also examined correlations between the week of epidemic onset and department level variables such as population size and elevation. For each department, we defined epidemic onset as the first week with reported cholera cases. Additionally, we generated maps of cholera attack rates by year. Maps were created using the choropleth package for R. All statistical analyses were performed using Matlab (The Mathworks, Inc).

2.3.1 Spatial Autocorrelation

Spatial autocorrelation is a measure of similarity of nearby observations. We assessed spatial autocorrelation of attack rates across departments using Moran's I statistic [48]. Moran's I is calculated using a nearest neighbor matrix w_{ij} of the 25 departments where $w_{ij} = 1$ when departments i and j share a border. All other entries are equal to zero. The statistic is calculated as in Equation 1, where N is the number of departments, x_i is the cholera incidence in department i, $\bar{x}$ is the mean cholera incidence across departments, and W is the sum of entries in matrix w_{ij}.

$$I = \frac{N \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{W \sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{1}$$

To determine significance of Moran's I, we used a nonparametric random data permutation test as in [49]. 10,000 random permutations of Peruvian departments were sampled given the observed data, generating a reference distribution of Moran's statistics under the null hypothesis of no spatial autocorrelation. P-values were calculated as the probability of obtaining the observed Moran's I or a more extreme value from the reference distribution [50].

2.3.2 Lorenz Curves and Gini Index

Finally, we quantified heterogeneity in attack rates using the Lorenz curve and Gini index [51, 52, 53, 54]. The Lorenz curve is a graphical representation showing the cumulative proportion of cholera cases plotted against the cumulative proportion of population. Under the assumption of homogeneity, the distributions will be balanced, and the Lorenz curve will fall on the 45 degree line. As heterogeneity of attack rates increases, the curve will become farther from this reference line. The Gini index is a summary measure of heterogeneity, calculated as the ratio of the area between the Lorenz curve and the reference line to the total area beneath the reference line. The value ranges from zero to one, with a larger value indicating greater heterogeneity.

2.4 Model Description and the Reproductive Number

We modeled the weekly trajectory of cholera incidence at two spatial scales: geographic region and department, using a standard compartmental dynamic model consisting of four equations (Equations 2-5) and 8 parameters (Table 1) that incorporates a variable transmission rate modulated by temperature [36, 37, 38]. In addition to the standard susceptible (S), infectious (I), and removed (R) compartments, this model includes a compartment (B) representing the concentration of vibrios in the environment. Hence, the model accounts for two transmission pathways: cholera exposure from the contaminated environment, and human to human transmission, likely corresponding to high dose infection and low dose infection, respectively [36]. In our model, we further incorporate the effect of weekly temperature variation ($Z(t)$) on the cholera transmission rate from the environment.

The corresponding compartmental model diagram is shown in Figure 2. Individuals in a population of size N are born and die at rate μ. Susceptible individuals can be infected through the environment with transmission rate $\beta_e(t)$ or through human contact with transmission rate β_h. Therefore, they move from susceptible to infectious classes at rates $\beta_e(t)\frac{B}{B+\kappa}$ (where κ is the 50% infectious dose in the environment and B is the current concentration of vibrios in the environment) and $\beta_h I$ respectively. Vibrios are shed by infectious individuals into the environment at rate λ and then die at rate δ. Infected individuals are assumed to recover at rate γ. The overall transmission dynamics can be mathematically described by the following set of nonlinear differential equations:

$$\frac{dS}{dt} = \mu N - \beta_h SI - \beta_e(t) S \frac{B}{B+\kappa} - \mu S \tag{2}$$

$$\frac{dI}{dt} = \beta_h SI + \beta_e(t) S \frac{B}{B+\kappa} - \mu I - \gamma I \tag{3}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{4}$$

$$\frac{dB}{dt} = \lambda I - \delta B \tag{5}$$

where $\beta_e(t) = \beta_{e1} + \beta_{e2} * Z(t)$

In our model, we further break down the parameter $\beta_e(t)$ into two components: $\beta_e(t) = \beta_{e1} + \beta_{e2} * Z(t)$, where $Z(t)$ represents the departments weekly standardized temperature at time t. We estimate three parameters (β_{e1}, β_{e2}, and β_h), as well as the initial concentration of vibrios in the environment, $B(0)$, and a reporting rate which is allowed to vary across departments. The reporting rate is simply a scaling factor used to adjust for the underreporting of cases, owing to, for instance, the large proportion of asymptomatic cholera cases. For simplicity, the department-specific reporting rate is not included in the model specification. Table 1 includes all model parameters and their definitions, as well as the values chosen for those that are fixed.

The basic reproductive number (R_0) is the most common measure of transmission potential

during the early phase of an epidemic [55]. R_0 is interpreted as the average number of new cases generated by a single infected individual in a completely susceptible population. Mathematically, this value assumes early exponential growth, and is therefore less useful during later generations of disease transmission [56, 57]. In our context of cholera transmission comprising multiple epidemic waves, we sought to monitor the time-dependent changes in the transmission potential associated with temperature fluctuations. For this purpose, the effective reproductive number, a time dependent measure of transmission denoted by $R(t)$ [58], can be used to describe the average number of secondary infections as the epidemic evolves. $R(t)$ can be derived by including the time-dependent transmission rate in the basic reproductive number. The basic reproductive number is derived from the associated next generation matrices following the method of van den Driessche and Watmough [56]. $R(t)$ reflects the changing transmission potential over time in the absence of susceptible depletion. Our model specification allows the definition of both an environmental component ($R_e(t)$) as well as a time-invariant human transmission component (R_h). These quantities can be estimated separately. We report two estimates for $R_e(t)$: a minimum value ($R_e(1)$) and a maximum value ($R_e(2)$) based on the observed ranges of temperature for each department from 1991-1993. Graphs are also included showing the effective environmental reproductive number over time. The formulas for each reproductive number are shown in Equations 6-8 [36].

$$R_t = \frac{N}{\delta\kappa(\gamma+\mu)}(\lambda\beta_e(t) + \delta\kappa\beta_h) \tag{6}$$

$$R_e = \frac{N\lambda\beta_e(t)}{\delta\kappa(\gamma+\mu)} \tag{7}$$

$$R_h = \frac{N\beta_h}{\gamma+\mu} \tag{8}$$

2.5 Parameter Estimation

We independently fit the model (Equations 2-5) to 24 departments and all three geographic regions for the first three years of the epidemic. We estimated five parameters (β_{e1}, β_{e2}, β_h, $B(0)$, and area-specific reporting rate) using approximate Bayesian computation (ABC) with a noninformative prior. To ensure the parameter space was sampled effectively, 100,000 estimates were randomly chosen for each of five parameters using Latin hypercube sampling. The cholera model was evaluated at each of these estimates, and we calculated the sum of squares error between model-estimated case incidence and true case counts. Parameter estimates minimized model error for each region and department.

By solving the system of ordinary differential equations at the resulting parameter estimates, a best fit incidence curve was generated for each set of best fit estimates. The root mean square error (RMSE) was evaluated for each model, though RMSE does not allow an objective comparison of model fit across departments. Therefore, in order to compare model fits, plots of observed versus predicted case counts are provided for each region and department. We also assessed R^2 for the

linear association between observed and predicted cases.

3 Results

Though it is possible that cholera appeared months earlier as previously suggested [29], the first cholera cases were reported to the Peruvian Ministry of Health in January of 1991. Infection spread so rapidly that by late February, the coastal departments had already received the brunt of the epidemic, though infection would reappear in subsequent years. As seen in Figure 3, the epidemic progressed in three waves, first hitting the coast where potential exposure to an aquatic reservoir was most likely, and subsequently spreading through the highlands and jungle regions. Along with the explosive onset and rapid progression of disease, this provides strong evidence for an environmental source of *V. cholerae*.

It appears that large populations along the coast were seeded first, with the epidemic thereafter spreading throughout the regions of Peru in a hierarchical manner. The epidemic peaked first in the coastal departments, with the highlands and then jungle regions following in the next few weeks. Throughout 1991, coastal departments saw far more cases than the remaining regions (Figure 4). However, this variability was likely a result of population size and density, because a different picture emerges when we examine the attack rates in each region (Figure 5a). Clearly, some of the highest attack rates occurred in the jungle. Loreto and Ucayali suffered attack rates as high as 3,000 per 100,000 persons in 1991. Yet, these departments were also among the last to be hit by the epidemic (Figure 5b).

Table 2 and Figure 6 show the progression of department level attack rates from 1991 to 1997. Not only was there significant variability in attack rates across regions, but there was also extreme variation within regions. For example, Moquegua saw an attack rate of less than 500 per 100,000 in 1991, while La Libertad saw over 2,600 cases per 100,000. Both are coastal departments. On the other hand, highlands departments appear to have had consitently low attack rates over the years.

Additionally, as seen in Figure 7a, we confirmed that departments with large populations tended to have earlier epidemic onsets ($\rho = -0.519, P < 0.01$). However, these departments were likely to be closer to the coast, which may have contributed to this relationship. On the other hand, there was no significant relationship between elevation and epidemic onset ($\rho = 0.212, P = 0.309$) (Figure 7b). Elevation did contribute to case incidence, likely by way of temperature, but it is most likely a combination of population size, density, and proximity to the coast that determined epidemic onset. Figure 8 looks more closely at the relationship between elevation and one-year cumulative incidence. In 1991, elevation seemed to provide significant protection against cholera ($\rho = -0.596, P < 0.01$). Thereafter, the relationship is less obvious, possibly because cholera had already established itself in all three regions. The trend over all seven years of the epidemic was also significant ($P < 0.01$) (Figure 9).

Spatial analysis revealed that Moran's I remained nonsignificant over the course of the epidemic (Figure 10). The P-value for all seven years of the outbreak was 0.08, perhaps indicating a lack of spatial autocorrelation, though the borderline significant P-values suggest it may be useful to further investigate correlations in case incidence between departments. From these results, however, we hypothesize that cholera incidence in one department was not strongly correlated with incidence in neighboring departments. Under the assumption that the epidemic was largely driven

by environmental contamination and not human to human transmission, the nonsignificant spatial autocorrelation is unsurprising.

Additionally, in 1991 attack rates were largely homogeneous in terms of population size ($G = -0.03$). However, as the epidemic progressed, there was increasing heterogeneity, with a Gini index close to -0.3 in both 1992 and 1995 (Figure 11). This indicates that at the onset of the epidemic, population size did not necessarily determine the burden of disease, though as the epidemic progressed, there was more significant association between cholera incidence and population size. The Gini index summarizing the entire epidemic (1991-1997) indicates mild heterogeneity of attack rates ($G = -0.16$).

As expected, there was also a clear seasonal trend throughout the course of the epidemic. Cases surged at the beginning of each year, a clear pattern emerging that would continue through 1995. We hypothesized that this seasonal trend was a result of several environmental factors, and we assessed how fluctuations in temperature might contribute to cholera incidence in each department. Averaging both minimum temperature and case incidence by week over the first three years of the epidemic, we saw strong correlations between the two variables, most clearly in the coastal departments (Figures 12, 13). The highlands region showed weaker positive correlations, two of which are examined more closely along with two coastal departments in Figure 14, whereas the jungle departments did not show a consistent relationship between temperature and cholera incidence. Though as temperature rises the warm coastal waters become an ideal reservoir for *V. cholerae*, temperature may have a smaller effect as we move farther from the coast, and human contact may become an important driver of transmission.

Finally, the compartmental model was used to estimate reproductive numbers, reporting rates, and initial concentrations of vibrios in the environment. We modelled time series by geographic region as well as for all 24 departments with sufficient case counts from 1991-1993. Model fit was assessed by evaluating the linear relationship between observed and predicted cases. Model fit statistics are presented in Table 3, along with statistics for model fit using fixed parameters, which is described further below. At the regional level, the model does not provide a strong fit to either the coastal, highland, or jungle departments ($0.4 < R^2 < 0.6$) (Figure 15). At the department level, though there are several poorly fit models, there are also several departments which have strong predictive ability. Most of the poor fits correspond to departments with unusually high or low case counts in comparison to the rest of Peru. Because the model performs better at the department level, indicating the need to account for heterogeneity at the regional level, parameter estimates are reported only at the department level.

In general, the model consistently captures the temporal waves of cholera transmission by incorporating fluctuations in weekly temperature (Figures 16, 17). This allowed estimation of a range of environmental reproductive numbers ($R_{e(1)}$ and $R_{e(2)}$), which are reported in Table 4 along with estimates of initial concentration and reporting rates separated by geographic region. Environmental effective reproductive numbers are plotted in Figure 18, showing the characteristic seasonal fluctuation in transmission potential. In most departments, the environmental reproductive numbers exceed the estimates for R_h, though this is not always the case. This suggests that environmental exposure played a larger role in transmission than did human contact. Interestingly, most departments in

which this did not hold true were grouped in the highlands and jungle regions, where, as previously mentioned, there was a dampened effect of temperature as compared to coastal departments.

There was also a wide range of estimates for the initial concentration of vibrios in the environment. Most estimates fell below 100,000 cells/mL, though several departments had initial concentrations well above this value. Several of these departments also had higher estimates for environmental reproductive numbers, confirming that the prevalence of vibrios in the environment will directly affect transmission potential.

Finally, estimates of reporting rates varied greatly, ranging from 0 to 35%. These estimates seem extremely low, though when examined in context, they are consistent with the ratio of symptomatic to asymptomatic cholera cases. Given that reporting rates are low and data is limited during such an epidemic, the performance of the model was examined using fixed parameter estimates while allowing only temperature to fluctuate. Parameter values were chosen by fitting the model to aggregated case counts for the country as a whole. Figure 19 shows a comparison of scaled data, calculated as proportion of total cases, and scaled model-estimated case counts by department. Though for unscaled case counts the model can perform only as well as the parameters we input, the scaled graphs show that it is possible to recover the temporal waves of transmission even in the absence of unique parameter estimates. Values for R^2 drop substantially when using fixed parameters, but the departments for which the original model provided a strong fit continue to reflect moderate relationships between observed and predicted cases (Table 3).

4 Discussion

Given the high values of environmental reproductive numbers, our results support the hypothesis that environmental transmission was the primary driver of the cholera epidemic in Peru. This study has characterized both the spatial dynamics of transmission as well as the temporal progression of the epidemic, thereby revealing important consequences of proximity to a large cholera reservoir. In Peru, initial transmission was characterized by multiple spatial waves. With the coastal waters serving as an aquatic reservoir, cholera spread first to the coastal departments before generating second and third waves that spread through the highlands and jungle, respectively. Each geographic region experienced different epidemic onsets and peak timing. Incidence peaked in coastal departments only a few weeks after the epidemic onset. However, the second and third waves of transmission had staggered onsets and epidemic peaks. These differences are likely a result of proximity to the aquatic reservoir and relative contributions of the two transmission routes: environment to human and human to human transmissions.

Despite the differences in transmission across regions, we did not identify significant spatial autocorrelation throughout the epidemic. However, statistical significance was borderline, and though we felt it was unnecessary to further analyze the data using a metapopulation approach, this may be useful for further research, as it is possible a different approach would lead to new insights. In this paper, we simply focused on department level transmission dynamics to characterize transmission parameters and reporting rates. We did, however, identify a potential association between population size and attack rates. This heterogeneity in case incidence may indicate that because more densely populated cities are closer to the coast, individuals may be more likely to come into contact with contaminated water.

Interestingly, all R_h estimates were below one, indicating that human to human transmission alone would not have sustained the epidemic. Conversely, R_e estimates were almost all greater than their respective R_h estimates, confirming the greater importance of waterborne transmission. These results contrast those of Mukandavire et al. in both Haiti and Zimbabwe [36, 37]. Transmission was driven by environmental contamination in Haiti, though the results were less consistent within departments [37]. Only one department clearly favored environmental transmission, though this department housed the contaminated river, resulting in an extremely high estimate of R_e. This weighted the contribution of environmental transmission in the countrys overall reproductive number. It should also be noted that these studies estimated only the time-invariant basic reproduction number characterizing early epidemic growth.

In Zimbabwe, Mukandavire et al. revealed that human to human by far outweighed environmental transmission [36]. This finding is very different than what we identified in Peru. It is clear that in Peru, transmission originated along the coast from a large aquatic reservoir recently upset by environmental processes. From there, cholera spread inland, but the coastal waters remained the primary driver of transmission. Alternatively, in a landlocked country such as Zimbabwe, human to human transmission may be more important. Additionally, the surprising result in Haiti could be a consequence of the location of contaminated water. The original contaminated source was found within the country rather than surrounding it.

Department estimates of the basic reproductive number range from 0.35 to 10.94, which is consistent with the cholera modeling literature [38, 59, 60, 61, 62], though our study is the first to use a model incorporating weekly fluctuations in temperature. Previous estimates have most often fallen between one and four, though in India and Europe some estimates have far exceeded ten [63, 39]. From this analysis, it is clear that Lima was the only department in Peru with such a large reproductive number. We would expect this to be a result of population size and density, though it was the R_e estimate and not the R_h estimate that was unusually high. It warrants further research to examine potential factors that inflate the environmental reproductive number besides temperature fluctuations.

It is also important to note that by incorporating temperature into the cholera model, it was possible to reproduce the shape of the incidence curve without estimating any model parameters. This may not be useful for predicting exact case counts, but it allows prediction of the broader temporal course of an epidemic by looking at temperature or climate. This will be more useful in settings like Peru, where environmental transmission is a clear driver of the epidemic. Additionally, in countries where cholera is endemic, analysis of temperature could indicate when prevention efforts should be intensified.

Additionally, these results are consistent with previous evidence that environmental factors contribute to cholera transmission patterns. For example, in the 1991-1996 epidemic in Mexico, Borroto et al. revealed a geographic pattern similar to that of Peru, demonstrating that coastal cities had much higher attack rates than did interior cities [64]. Other studies have also confirmed that proximity to a water source results in higher incidence of cholera infection [65, 66]. Similarly, complex environmental factors have also been implicated in cholera transmission patterns. For example, Nkoko et al. studied the Great Lakes Region of Africa from 1978 to 2008 and found that abnormally warm El Niño events corresponded to increases in cholera incidence [67]. This is consistent with hypotheses that El Niño may have influenced the 1991 epidemic in Peru [29].

Ngwa et al. used a regression model to identify associations between risk of cholera transmission and environmental variables in Cameroon [68]. They found significant associations with both average daily maximum temperatures and precipitation levels. However, the current study may be the first to examine time series correlations between temperature and cholera incidence. Though it is well established that temperature plays an important role in the persistence of *V. cholerae*, the aforementioned results show that fluctuations in temperature closely resemble temporal patterns of cholera incidence. It is important to note, though, that this relationship is stronger in geographic regions closest to an aquatic reservoir. In addition, though we did not find a strong effect of temperature in the jungle region of Peru, these departments had some of the highest attack rates in 1991. This may be related to the accessibility of clean water and sanitation services.

Though these findings should be useful for further research, the results are still subject to limitations. As mentioned previously, cholera cases are often underreported. Though we attempted to correct for this, it is possible that parameter estimates are biased. On the other hand, the broad definition for suspected cholera infection may include non-cholera cases [46]. Additionally, it was impossible to appropriately fit the model to every department in Peru given the small number of cases in some areas. Finally, we were unable to produce confidence intervals using ABC methods,

and so only point estimates have been reported. Ideally, ABC would produce a posterior distribution useful for parameter inference.

5 Conclusions

By incorporating environmental variables into a cholera epidemic model, we were able to effectively describe multiple waves of transmission characteristic of *V. cholerae*. This allows for more accurate estimation of model parameters and could be useful in future studies which predict the course of an epidemic. Additionally, the model effectively differentiates transmission patterns by geographic region even in the absence of unique parameter estimates. The effect of temperature alone recreates the shape of transmission waves, so with a single set of parameters, fluctuations in cholera incidence can be approximated in geographic regions with varying climates. These results indicate that mathematical models can provide valuable information about transmission patterns and should continue to be used to inform public health decision making.

6 Bibliography

[1] Mohammad Ali, Allyson R. Nelson, Anna Lena Lopez, and David A. Sack. Updated global burden of cholera in endemic countries. *PLoS Neglected Tropical Diseases*, 9(6):e0003832, 2015.

[2] David A. Sack, R. Bradley Sack, G. Balakrish Nair, and A. K. Siddique. Cholera. *The Lancet*, 363(9404):223–233, 2004.

[3] Rita R. Colwell. Global climate and infectious disease: The cholera paradigm. *Science*, 274(5295):2025, 1996.

[4] Tatsuo Kaneko and Rita R. Colwell. Ecology of vibrio parahaemolyticus in chesapeake bay. *Journal of Bacteriology*, 113(1):24–32, 1973.

[5] James D. Oliver, Robert A. Warner, and David R. Cleland. Distribution of vibrio vulnificus and other lactose-fermenting vibrios in the marine environment. *Applied and Environmental Microbiology*, 45(3):985–998, 1983.

[6] G. Balakrish Nair, Firdausi Qadri, Jan Holmgren, Ann-Mari Svennerholm, Ashrafus Safa, Nurul A. Bhuiyan, Q. Shafi Ahmad, Shah M. Faruque, A. S. G. Faruque, Yoshifumi Takeda, and David A. Sack. Cholera due to altered el tor strains of vibrio cholerae o1 in bangladesh. *Journal of Clinical Microbiology*, 44(11):4211–4213, 2006.

[7] Jr Ira M. Longini, Mohammed Yunus, K. Zaman, A. K. Siddique, R. Bradley Sack, and Azhar Nizam. Epidemic and endemic cholera trends over a 33-year period in bangladesh. *The Journal of Infectious Diseases*, 186(2):246–251, 2002.

[8] Anna Lena Lopez, Maria Liza Antoinette Gonzales, Josephine G. Aldaba, and G. Balakrish Nair. Killed oral cholera vaccines: history, development and implementation challenges. *Therapeutic Advances in Vaccines*, 2(5):123–136, 2014.

[9] J. L. Sanchez, B. Vasquez, R. E. Begue, R. Meza, G. Castellares, C. Cabezas, D. M. Watts, A. M. Svennerholm, J. C. Sadoff, and D. N. Taylor. Protective efficacy of oral whole-cell/recombinant-b-subunit cholera vaccine in peruvian military recruits. *The Lancet*, 344(8932):1273–1276.

[10] David N. Taylor, Vicky Crdenas, Jos L. Sanchez, Rodolfo E. Bgu, Robert Gilman, Christian Bautista, Juan Perez, Romulo Puga, Alvaro Gaillour, Rina Meza, Peter Echeverria, and Jerald Sadoff. Two-year study of the protective efficacy of the oral whole cell plus recombinant b subunit cholera vaccine in peru. *The Journal of Infectious Diseases*, 181(5):1667–1673, 2000.

[11] E. Gotuzzo. El clera en el per. *Revista Mdica Herediana*, 2(3), 1991.

[12] R. Finkelstein. *Cholera, Vibrio cholerae O1 and O139, and Other Pathogenic Vibrios*, book section 24. University of Texas Medical Branch at Galveston, Galveston, TX, 4 edition, 1996.

[13] W. E. Woodward and W. H. Mosley. The spectrum of cholera in rural bangladesh. ii. comparison of el tor ogawa and classical inaba infection. *Am J Epidemiol*, 96(5):342–51, 1972.

[14] C. Carrillo. El instituto nacional de salud y la epidemia de clera. *Revista Mdica Herediana*, 2(2), 1991.

[15] Jason B. Harris, Regina C. LaRocque, Firdausi Qadri, Edward T. Ryan, and Stephen B. Calderwood. Cholera. *The Lancet*, 379(9835):2466–2476, 2012.

[16] Ankur Mutreja, Dong Wook Kim, Nicholas Thomson, Thomas R. Connor, Je Hee Lee, Samuel Kariuki, Nicholas J. Croucher, Seon Young Choi, Simon R. Harris, Michael Lebens, Swapan Kumar Niyogi, Eun Jin Kim, T. Ramamurthy, Jongsik Chun, James L. N. Wood, John D. Clemens, Cecil Czerkinsky, G. Balakrish Nair, Jan Holmgren, Julian Parkhill, and Gordon Dougan. Evidence for multiple waves of global transmission within the seventh cholera pandemic. *Nature*, 477(7365):462–465, 2011.

[17] D. Barua. The global epidemiology of cholera in recent years. *Proceedings of the Royal Society of Medicine*, 65(5):423–428, 1972.

[18] Asish K. Mukhopadhyay, Yoshifumi Takeda, and G. Balakrish Nair. *Cholera Outbreaks in the El Tor Biotype Era and the Impact of the New El Tor Variants*, pages 17–47. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[19] R. W. Goodgame, W. B. Greenough, and Iii. Cholera in africa: A message for the west. *Annals of Internal Medicine*, 82(1):101–106, 1975.

[20] H. Kstner, I. Gibson, T. Carmichael, L. Van Zyl, C. Chouler, J. Hyde, and J. Du Plessis. The spread of cholera in south africa. *South African Medical Journal*, 60:87–90, 1981.

[21] WilliamB Baine, Mirella Mazzotti, Donato Greco, Egidio Izzo, Alfredo Zampieri, Guiseppe Angioni, Mario Di Gioia, EugeneJ Gangarosa, and Francesco Pocchiari. Epidemiology of cholera in italy in 1973. *The Lancet*, 304(7893):1370–1374, 1974.

[22] Cholera in spain. *British Medical Journal*, 3(5769):266–266, 1971.

[23] O. Weil and P. Berche. *The cholera epidemic in Ecuador: Towards an endemic in Latin America*, volume 40. 1992.

[24] L. Seminario, A. Lpez, E. Vsquez, and M. Rodriguez. Epidemia de clera en el per: vigilancia epidemiolgica. *Revista Peruana de Epidemiologa*, 4(2):8–41, 1991.

[25] H. Lores. Situacin del clera en las americas- papel de la organizacin panamericana de la salud. *Revista Peruana de Epidemiologa*, 4(2):109–110, 1991.

[26] J. P. Guthmann. Epidemic cholera in latin america: spread and routes of transmission. *The Journal of tropical medicine and hygiene*, 98(6):419–427, 1995.

[27] Cholera in the americas. *Bulletin of the Pan American Health Organization*, 25(3):267–273, 1991.

[28] M. Cuadra, J. Neyra, and A. Cuadra. El clera en un pas desarrollado y en un subdesarrollado. el clera en el per visto en la perspectiva de un reciente caso en berln, alemania. *Revista Peruana de Epidemiologa*, 6(2):29–33, 1993.

[29] C Seas, J Miranda, A I Gil, R Leon-Barua, J Patz, A Huq, R R Colwell, and R B Sack. New insights on the emergence of cholera in latin america during 1991: the peruvian experience. *The American Journal of Tropical Medicine and Hygiene*, 62(4):513–517, 2000.

[30] B. Benavides. Consumo de alimentos en comedores populares de lima durante la epidemia de clera. *Revista Peruana de Epidemiologa*, 4(2):102–108, 1991.

[31] J. Cieza, G. Gamara, and C. Torres. Letalidad y riesgo de insuficiencia renal por clera en el hospital nacional cayetano heredia de lima, per. *Revista Mdica Herediana*, 2(2), 1991.

[32] C. Gastaaga Ruiz. Agua y saneamiento bsico, per 1991. *Revista Peruana de Epidemiologa*, 4(2):70–85, 1991.

[33] M. Rodrguez, E. Tejada, L. Seminario, D. L. Swerdlow, E. D. Mintz Blake, K. Greene, J. G. Wells, C. Ocampo, L. Espejo, W. Saldaa, M. Pollack, and R. V. Tauke. Epidemia del clera en el distrito de vctor larco herrera, trujillo. la libertad, per. *Revista Peruana de Epidemiologa*, 4(2):42–46, 1991.

[34] J. Uribe, Ochoa, H. Ortiz Souza, G. Pemberton Medina, and W. Ortiz Alvarez. Clera en el per: Primer caso diagnosticado clnicamente. *Revista de Farmacologa y Teraputica*, 1(1):8–9, 1991.

[35] M. Yabar, F. Torres, J. Tapia, and M. Picn. Situacin del clera en la provincia de cajamarca. *Revista Peruana de Epidemiologa*, 6(1):23–26, 1993.

[36] Zindoga Mukandavire, Shu Liao, Jin Wang, Holly Gaff, David L. Smith, and J. Glenn Morris. Estimating the reproductive numbers for the 20082009 cholera outbreaks in zimbabwe. *Proceedings of the National Academy of Sciences of the United States of America*, 108(21):8767–8772, 2011.

[37] Zindoga Mukandavire, David L. Smith, and J. Glenn Morris Jr. Cholera in haiti: Reproductive numbers and vaccination coverage estimates. *Scientific Reports*, 3:997, 2013.

[38] Andrew S. Azman, Francisco J. Luquero, Amabelia Rodrigues, Pedro Pablo Palma, Rebecca F. Grais, Cunhate Na Banga, Bryan T. Grenfell, and Justin Lessler. Urban cholera transmission hotspots and their implications for reactive vaccination: Evidence from bissau city, guinea bissau. *PLOS Neglected Tropical Diseases*, 6(11):e1901, 2012.

[39] Christina H. Chan, Ashleigh R. Tuite, and David N. Fisman. Historical epidemiology of the second cholera pandemic: Relevance to present day disease dynamics. *PLOS ONE*, 8(8):e72498, 2013.

[40] Experiencia de la epidemia del clera en el per 1991. In *FAO/WHO Global Forum of Food Safety Regulators*.

[41] D. Gonzalez del Carpio. El hospital bajo la furia del clera. *Revista Mdica Herediana*, 2(2), 1991.

[42] L. Seminario C, O. M. Legua, and D. Fishbein. Priorities for public health surveillance when resources are limited. *Morbidity and Mortality Weekly Report*, 41:85–89, 1992.

[43] 2018.

[44] D. Koo, H. Traverso, M. Libel, C. Drasbek, R. Tauxe, and D. Brandling-Bennett. Epidemic cholera in latin america, 1991-1993: Implications of case definitions used for public health surveillance. *Bulletin of the Pan American Health Organization*, 30(2):134–143, 1996.

[45] D. J. Vugia, J. E. Koehler, and A. A. Ries. Surveillance for epidemic cholera in the americas: An assessment. *Morbidity and Mortality Weekly Report*, 41(1):27–34, 1992.

[46] R. del Aguila, B. Benavides, E. Jacoby, and J. Novara. Reconocimiento de clera por personas sintomticas despus del brote epidmico en las udes lima sur y la sub-regin luciano castillo-regin grau. *Revista Peruana de Epidemiologa*, 5(1):5–9, 1992.

[47] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hlm, L. Isaksen, P. Kllberg, M. Khler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thpaut, and F. Vitart. The era-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011.

[48] A.D. Cliff and J.K. Ord. *Spatial Processes: Models & Applications.* Pion, 1981.

[49] Gerardo Chowell, Sherry Towers, Ccile Viboud, Rodrigo Fuentes, Viviana Sotomayor, Lone Simonsen, Mark A. Miller, Mauricio Lima, Claudia Villarroel, Monica Chiu, Jose E. Villarroel, and Andrea Olea. The influence of climatic conditions on the transmission dynamics of the 2009 a/h1n1 influenza pandemic in chile. *BMC Infectious Diseases*, 12(1):298, 2012.

[50] E. Edgington. *Randomization Tests, Fourth Edition.* Taylor & Francis, 1995.

[51] MEJ Woolhouse, C. Dye, J. F. Etard, T. Smith, JD Charlwood, GP Garnett, P. Hagan, JLK Hii, PD Ndhlovu, RJ Quinnell, CH Watts, SK Chandiwana, and RM Anderson. Heterogeneities in the transmission of infectious agents: Implications for the design of controlprograms. *Proceedings of the National Academy of Sciences of the United States of America*, 94(1):338–342, 1997.

[52] Roxanne P. Kerani, Mark S. Handcock, H. Hunter Handsfield, and King K. Holmes. Comparative geographic concentrations of 4 sexually transmitted infections. *American Journal of Public Health*, 95(2):324–330, 2005.

[53] Chris G. Green, Dennis O. Krause, and John L. Wylie. Spatial analysis of campylobacter infection in the canadian province of manitoba. *International Journal of Health Geographics*, 5:2–2, 2006.

[54] Gerardo Chowell, Lus M. A. Bettencourt, Niall Johnson, Wladimir J. Alonso, and Ccile Viboud. The 19181919 influenza pandemic in england and wales: spatial patterns in transmissibility and mortality impact. *Proceedings of the Royal Society B: Biological Sciences*, 275(1634):501–509, 2008.

[55] RM Anderson and RM May. Directly transmitted infections diseases: control by vaccination. *Science*, 215(4536):1053–1060, 1982.

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

[57] O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7(47):873–885, 2010.

[58] Hiroshi Nishiura and Gerardo Chowell. *The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends*, pages 103–121. Springer Netherlands, Dordrecht, 2009.

[59] A. R. Tuite, J. Tien, M. Eisenberg, D. D. Earn, J. Ma, and D. N. Fisman. 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, 2011.

[60] Dennis L. Chao, M. Elizabeth Halloran, and Ira M. Longini. Vaccination strategies for epidemic cholera in haiti with implications for the developing world. *Proceedings of the National Academy of Sciences*, 108(17):7081–7085, 2011.

[61] Jr. Longini, Ira M., Azhar Nizam, Mohammad Ali, Mohammad Yunus, Neeta Shenvi, and John D. Clemens. Controlling endemic cholera with oral vaccines. *PLOS Medicine*, 4(11):e336, 2007.

[62] Rachael L. Miller Neilan, Elsa Schaefer, Holly Gaff, K. Renee Fister, and Suzanne Lenhart. Modeling optimal intervention strategies for cholera. *Bulletin of Mathematical Biology*, 72(8):2004–2018, 2010.

[63] Rachael L. Miller Neilan, Elsa Schaefer, Holly Gaff, K. Renee Fister, and Suzanne Lenhart. Modeling optimal intervention strategies for cholera. *Bulletin of Mathematical Biology*, 72(8):2004–2018, 2010.

[64] Ren J. Borroto and Ramon Martinez-Piedra. Geographical patterns of cholera in mexico, 19911996. *International Journal of Epidemiology*, 29(4):764–772, 2000.

[65] Godfrey Bwire, Mohammad Ali, David A. Sack, Anne Nakinsige, Martha Naigaga, Amanda K. Debes, Moise C. Ngwa, W. Abdullah Brooks, and Christopher Garimoi Orach. Identifying cholera "hotspots" in uganda: An analysis of cholera surveillance data from 2011 to 2016. *PLOS Neglected Tropical Diseases*, 11(12):e0006118, 2017.

[66] Frank B. Osei, Alfred A. Duker, Ellen-Wien Augustijn, and Alfred Stein. Spatial dependency of cholera prevalence on potential cholera reservoirs in an urban area, kumasi, ghana. *International Journal of Applied Earth Observation and Geoinformation*, 12(5):331–339, 2010.

[67] D. B. Nkoko, P. Giraudoux, P. Plisnier, A. M. Tinda, M. Piarroux, B. Sudre, S. Horion, J. M. Tamfum, B. K. Ilunga, and R. Piarroux. Dynamics of cholera outbreaks in great lakes region of africa, 19782008. *Emerging Infectious Diseases*, 7(11):2026–2034, 2011.

[68] Moise C. Ngwa, Song Liang, Ian T. Kracalik, Lillian Morris, Jason K. Blackburn, Leonard M. Mbam, Simon Franky Baonga Ba Pouth, Andrew Teboh, Yang Yang, Mouhaman Arabi, Jonathan D. Sugimoto, and Jr. Morris, John Glenn. Cholera in cameroon, 2000-2012: Spatial and temporal analysis at the operational (health district) and sub climate levels. *PLOS Neglected Tropical Diseases*, 10(11):e0005105, 2016.

[69] Cludia Torres Codeo. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC Infectious Diseases*, 1:1–1, 2001.

[70] T. R. Hendrix. The pathophysiology of cholera. *Bulletin of the New York Academy of Medicine*, 47(10):1169–1180, 1971.

[71] J. B. Kaper, J. G. Morris, and M. M. Levine. Cholera. *Clinical Microbiology Reviews*, 8(1):48–86, 1995.

7 Appendix I

Tables and Figures

Table 1: Parameter symbols, definitions, units, and baseline values and ranges for Model 1.

Parameter	Definition	Value	Source	Range
μ	Natural birth & death rates	$\frac{1}{60*365}\ days^{-1}$		
κ	50% infectious dose	10^6 cells/mL	[69]	9^6 - 11^6 cells/mL
γ	Recovery rate	$\frac{1}{5}\ days^{-1}$	[70]	$\frac{1}{4}$ - $\frac{1}{6}\ days^{-1}$
λ	Rate of contribution of vibrios from infected individuals to the environment	$10\ cells * mL^{-1} * day^{-1}$ per person	[69]	8 - 12 $cells * mL^{-1} * day^{-1}$ per person
δ	Death rate of vibrios in the environment	$\frac{1}{30}\ days^{-1}$	[71]	$\frac{1}{25}$ - $\frac{1}{35}\ days^{-1}$
β_h	Human to human transmission rate	Estimated		
β_e	Environment to human transmission rate	Estimated		

Table 2: Attack rates per 100,000 by department for years 1991-1997.

	1991	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Tumbes	1304.24	767.01	82.30	44.11	32.92	0	0
Piura	1697.82	573.74	82.87	102.40	46.82	5.31	1.29
Lambayeque	1993.48	931.68	109.79	30.40	58.09	27.15	12.66
La Libertad	2608.99	990.20	218.00	90.62	142.27	23.53	10.46
Ancash	2017.60	625.00	358.74	145.55	139.65	17.52	23.63
Lima	1563.51	3427.94	586.54	214.25	373.15	39.10	12.03
Callao	2431.90	3318.72	430.98	458.25	178.84	4.01	0
Ica	535.84	654.07	392.33	134.84	116.11	17.85	4.06
Arequipa	2509.18	3722.56	581.76	158.99	17.38	0.98	2.06
Moquegua	487.43	1059.35	772.22	48.51	22.69	11.74	5.48
Tacna	730.42	313.37	267.87	10.79	0.94	0	0
Coast	1772.93	2327.69	426.26	171.03	222.26	25.34	9.71
Apurimac	24.85	70.74	10.14	18.26	2.79	0	0
Ayacucho	629.55	492.65	490.94	312.87	50.33	43.28	0.19
Cajamarca	702.48	611.63	479.72	244.46	116.61	28.57	17.80
Cusco	507.95	73.96	30.67	37.88	1.52	1.23	0
Huancavelica	462.83	274.61	149.38	50.04	54.03	2.49	1.00
Huanuco	610.90	436.66	179.86	57.47	111.45	22.05	1.06
Junin	867.68	467.19	302.11	95.61	57.53	7.78	2.78
Pasco	169.42	118.84	105.27	28.78	53.05	3.70	1.23
Puno	22.09	15.77	46.39	134.39	22.00	2.11	1.74
Highlands	500.23	311.58	221.38	124.75	55.38	13.04	4.34
Amazonas	836.86	1528.25	616.92	136.19	263.69	120.54	0
Loreto	3824.30	914.44	758.62	108.39	219.47	68.17	5.93
Madre de Dios	14.08	1.56	0	65.71	0	1.56	0
San Martin	1865.88	1027.57	522.87	64.02	85.93	20.41	8.99
Ucayali	3149.86	612.77	878.94	278.92	652.34	284.48	64.42
Jungle	2533.44	976.59	663.33	126.43	251.29	95.99	14.66

Table 3: Root mean square error (RMSE) and R^2 for model fits with and without parameter estimation.

	Best Fit Estimates		**Fixed Parameters**	
	RMSE	R^2	RMSE	R^2
Tumbes	31.78	0.51	1.28	0.23
Piura	224.53	0.81	0.38	0.6
Lambayeque	308.17	0.64	0.49	0.44
La Libertad	361.96	0.79	0.42	0.55
Lima	2260.84	0.27	0.47	0.16
Callao	137.37	0.77	0.43	0.56
Ica	52.88	0.55	0.59	0.27
Arequipa	198.52	0.67	0.60	0.14
Moquegua	31.17	0.04	1.71	0.01
Tacna	16.73	0.38	1.42	0
Coast	3086.07	0.58	0.39	0.44
Apurimac	7.80	0.10	0.99	0
Ayacucho	44.55	0.19	0.75	0.03
Cajamarca	199.55	0.41	0.50	0.21
Cusco	45.71	0.43	0.54	0.28
Huancavelica	27.12	0.29	0.96	0
Huanuco	32.04	0.64	0.52	0.32
Junin	73.33	0.59	0.54	0.21
Pasco	11.87	0.23	1.38	0.05
Puno	10.99	0.02	0.60	0.04
Highlands	267.40	0.53	0.42	0.35
Amazonas	70.03	0.32	0.81	0.1
Loreto	213.49	0.62	0.63	0.06
Madre de Dios	0.42	0.11	1.94	0.01
San Martin	66.71	0.62	0.52	0.29
Ucayali	88.08	0.61	1.01	0.02
Jungle	408.51	0.44	0.47	0.12

Table 4: Parameter estimates for 24 Peruvian departments. Horizontal lines separate the coast, highland, and jungle regions.

	R_h	$R_e(1)$	$R_e(2)$	$B(0)$ (cells/mL)	Reporting rate (%)
Tumbes	0.43	0.51	2.65	104862	3.0
Piura	0.95	1.18	5.90	84492	3.4
Lambayeque	0.97	0.64	2.07	61039	7.4
La Libertad	0.90	0.49	7.89	97866	6.0
Lima	0.35	0.00	10.59	130678	11.2
Callao	0.53	0.00	2.47	114886	14.1
Ica	0.42	0.00	2.36	90214	3.6
Arequipa	0.57	0.48	0.91	81258	35.0
Moquegua	0.41	0.39	1.27	52687	4.6
Tacna	0.76	0.25	0.46	53446	6.7
Apurimac	0.67	0.31	0.72	69707	0.4
Ayacucho	0.32	0.63	1.07	58611	10.7
Cajamarca	0.94	0.00	3.60	90054	2.1
Cusco	0.47	1.54	3.73	300410	0.8
Huancavelica	0.73	0.31	0.58	60212	4.0
Huanuco	0.78	0.04	1.01	60749	3.2
Junin	0.71	0.32	1.17	121717	4.4
Pasco	0.49	0.00	1.46	69062	1.7
Puno	0.40	0.41	1.15	359611	0.7
Amazonas	0.40	0.00	1.32	90010	15.1
Loreto	0.89	0.32	0.79	59117	13.4
Madre de Dios	0.92	0.00	1.10	63518	0.0
San Martin	0.59	0.47	1.69	53173	5.3
Ucayali	0.94	0.21	0.52	50921	13.3

Figure 1: Weekly mean, minimum, and maximum temperatures by department, 1991-1997.

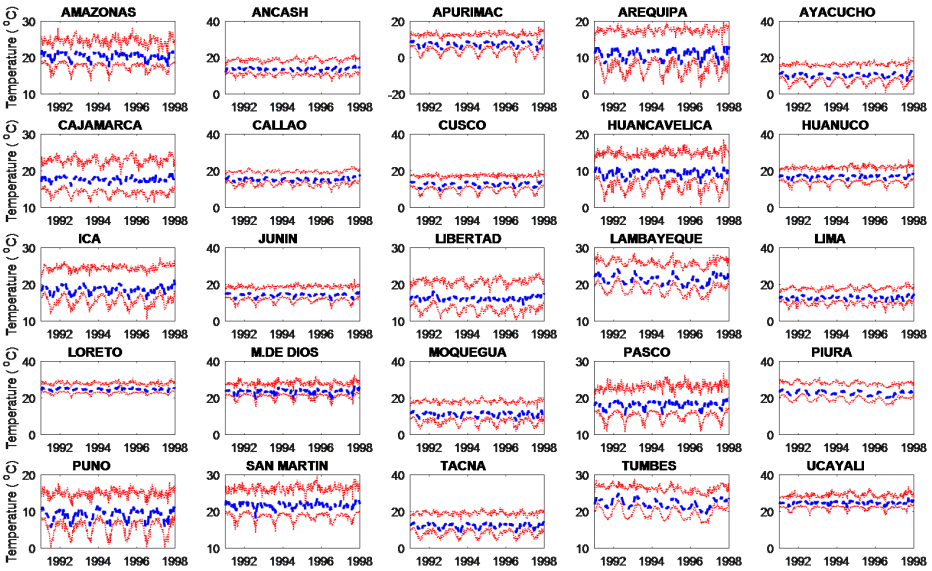


Figure 2: Model diagram. Susceptible individuals can be infected through the environment, where compartment B represents the current concentration of vibrios in the water supply, or through human contact. Infected individuals move to compartment R after dying or recovering.

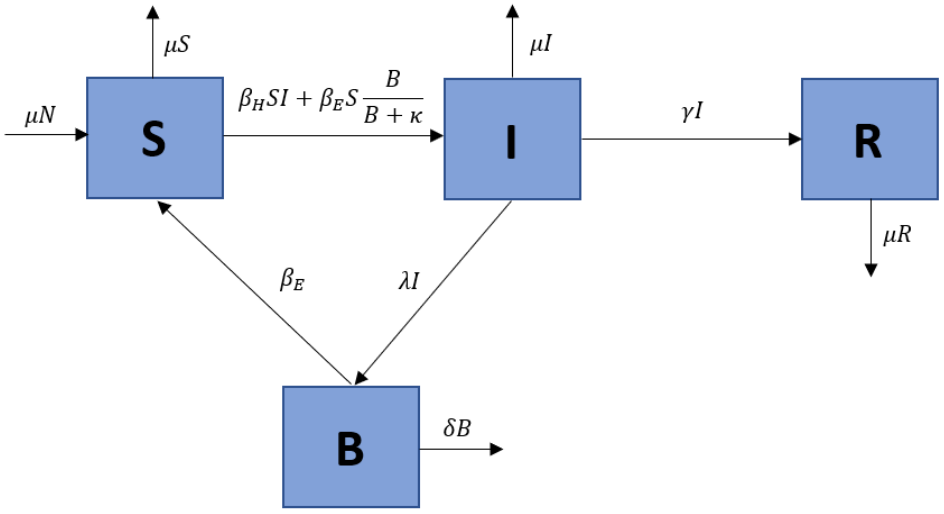


Figure 3: Weekly incidence of suspected cholera cases in Peru by region, January 1991 through December 1997. Curves represent the national and regional weekly proportions of total cases. The epidemic peaked first in the coastal departments and shortly thereafter in the highlands and jungle regions. Throughout the years, there was a clear seasonal trend of cholera infection with peaks occurring at the beginning of each year.

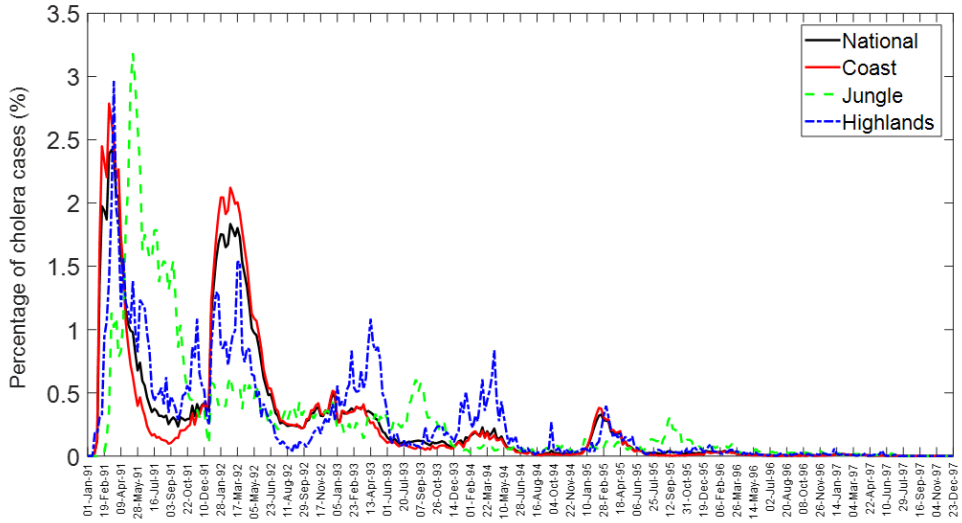


Figure 4: Color scale image of weekly cholera cases by department. Weekly cases have been log scaled, and dashed lines separate the coast, highland, and jungle regions. The epidemic hit the coastal departments early, and the highest case counts were concentrated in this region as well.

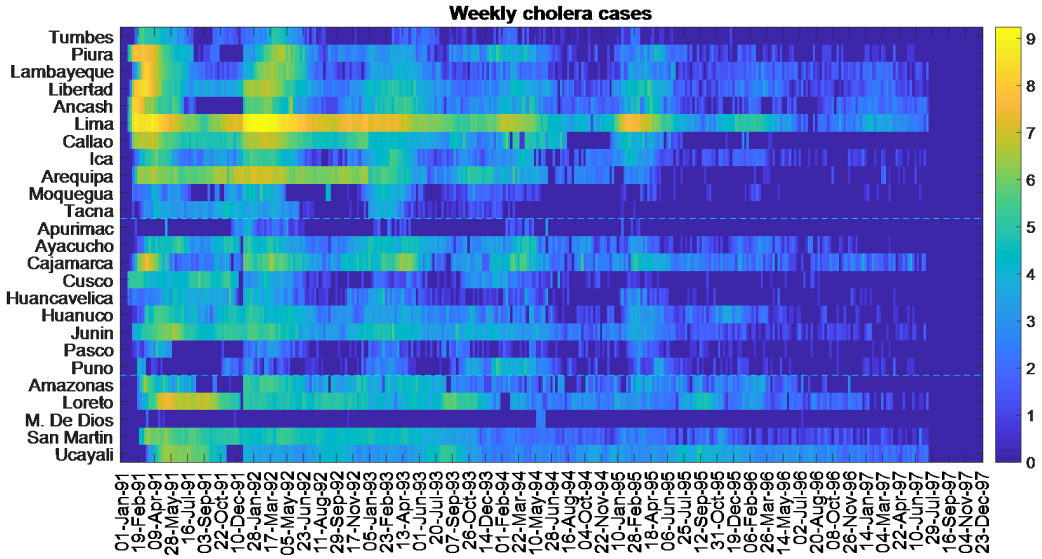


Figure 5: a. Attack rates of cholera per 100,000 for Peruvian departments in 1991. The highest attack rates occurred in the jungle region, specifically departments such as Loreto, Ucayali, and San Martin. Coastal departments also showed consistently high attack rates. b. Map showing week of epidemic onset by department. Darker regions experienced a later onset, defined as the first week in 1991 with reported cholera cases. Ucayali and Madre de Dios were the last to be hit by the epidemic.

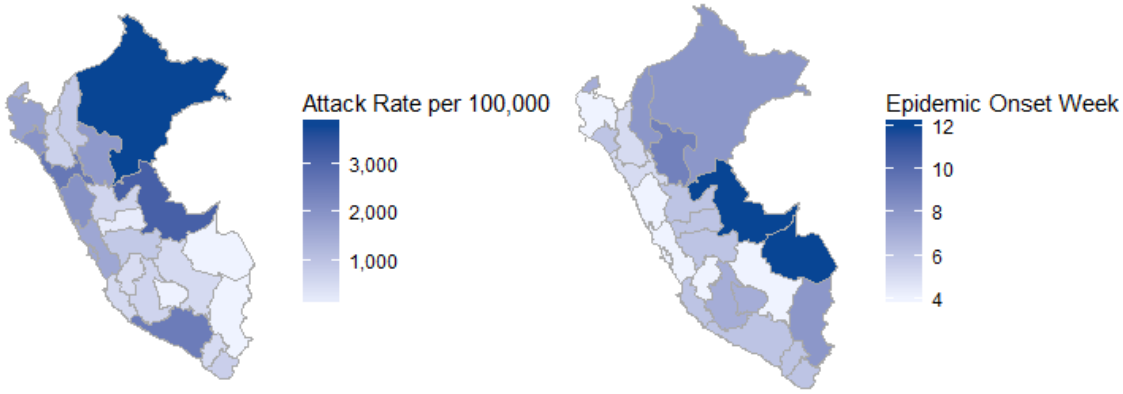


Figure 6: Temporal progression in attack rates by department and region for the years 1992-1997.

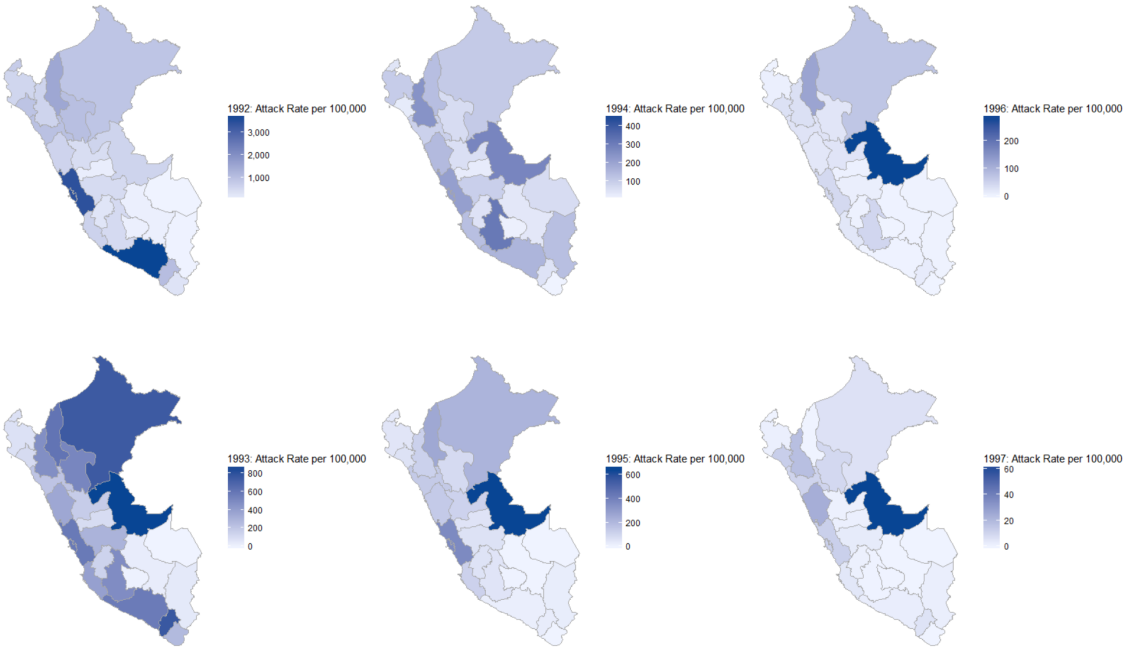


Figure 7: a. Correlation between department population and week of epidemic onset. Departments with larger populations tend to have an earlier epidemic onset ($P < 0.01$). b. Correlation between department elevation and week of epidemic onset. There is a non-significant trend for departments with higher elevation to have later epidemic onsets.

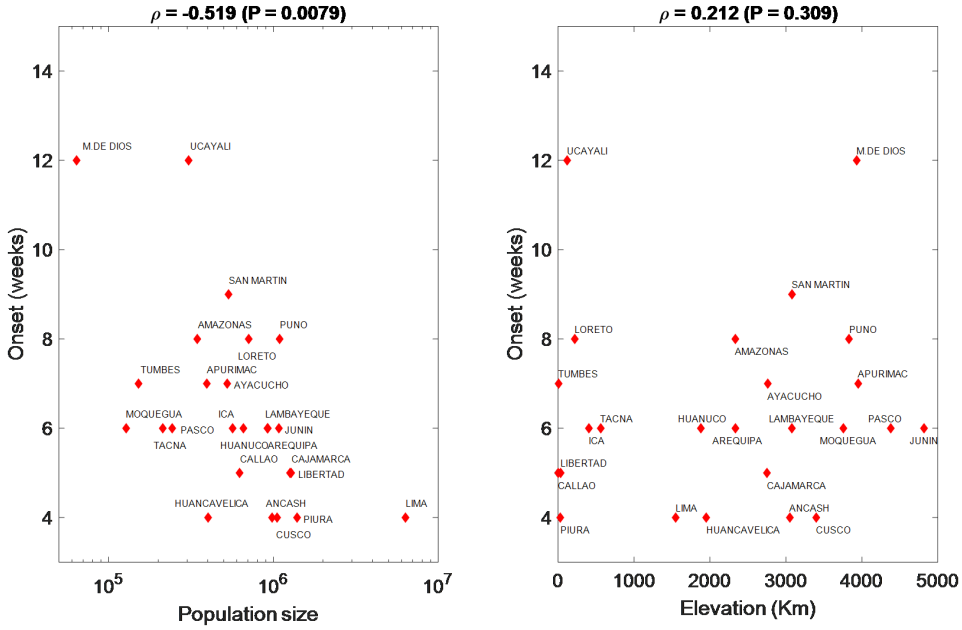


Figure 8: Correlations between department incidence rate and elevation by year. Departments with higher elevation tended to have lower one-year cumulative incidence rates.

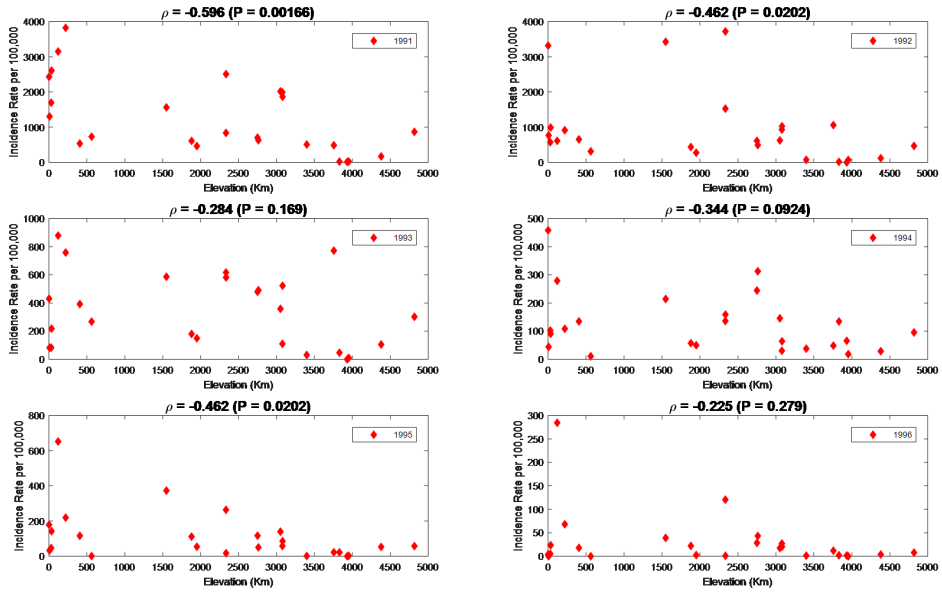


Figure 9: Correlation between department seven-year cumulative incidence rate and elevation. Departments with higher elevation had a lower incidence rate from 1991-1997.

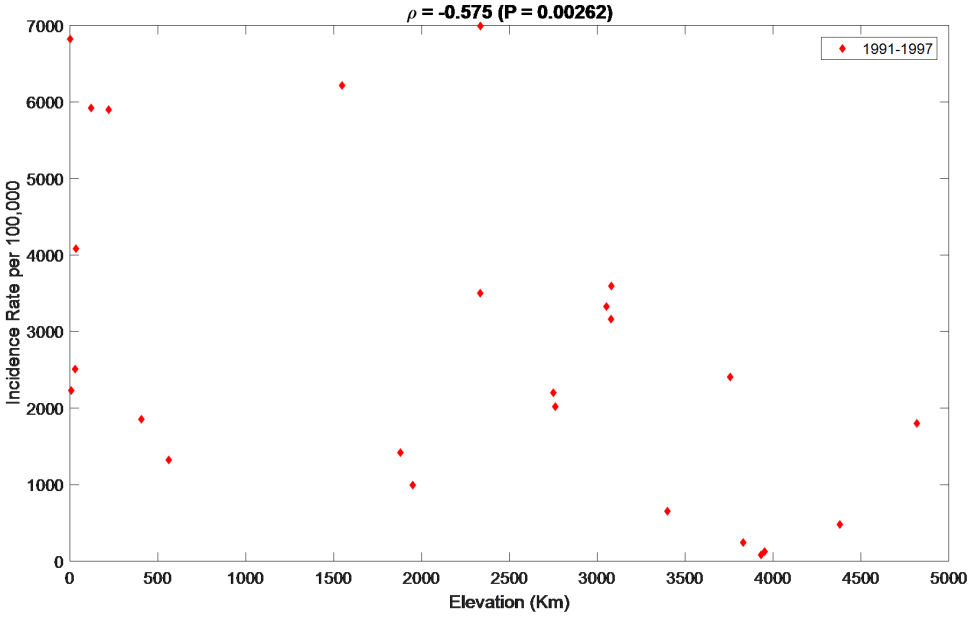


Figure 10: Significance of Morans I by year, 1991-1997. There was minimal evidence of spatial autocorrelation for all years of the epidemic. The P-value for all seven years combined was nonsignificant at 0.08.

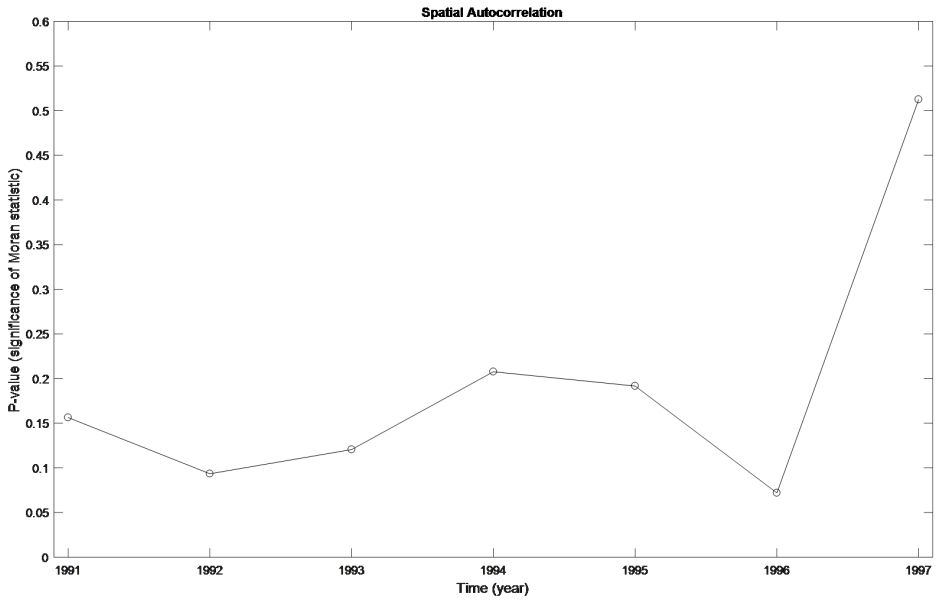


Figure 11: Lorenz curves assessing heterogeneity by year. Attack rates were mostly homogeneous, though there is indication of mild heterogeneity in both 1992 and 1995.

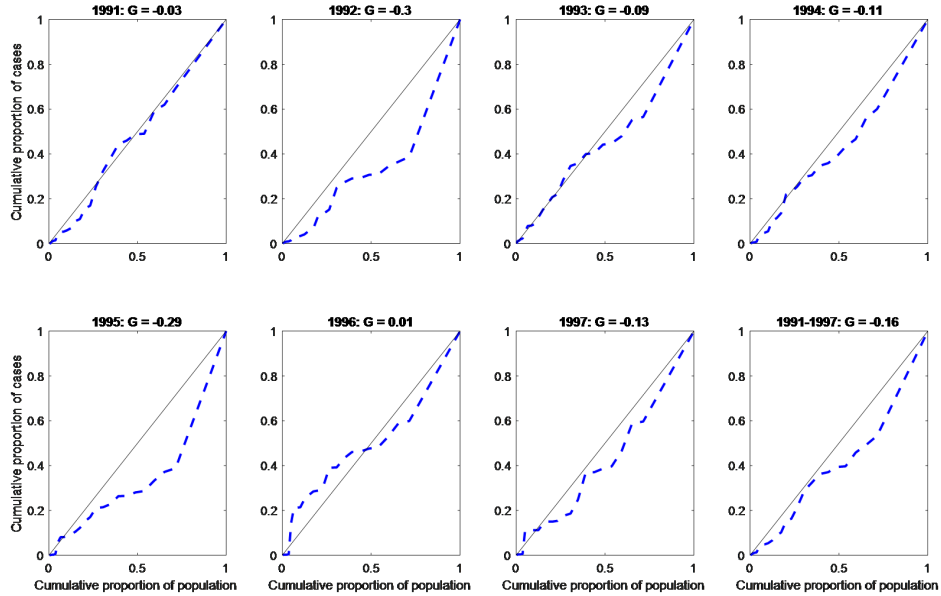


Figure 12: Average weekly cases and minimum temperatures by department are calculated by taking the mean over the first three epidemic years. Mean scaled incidence and mean minimum temperature are plotted over a 52 week time period.

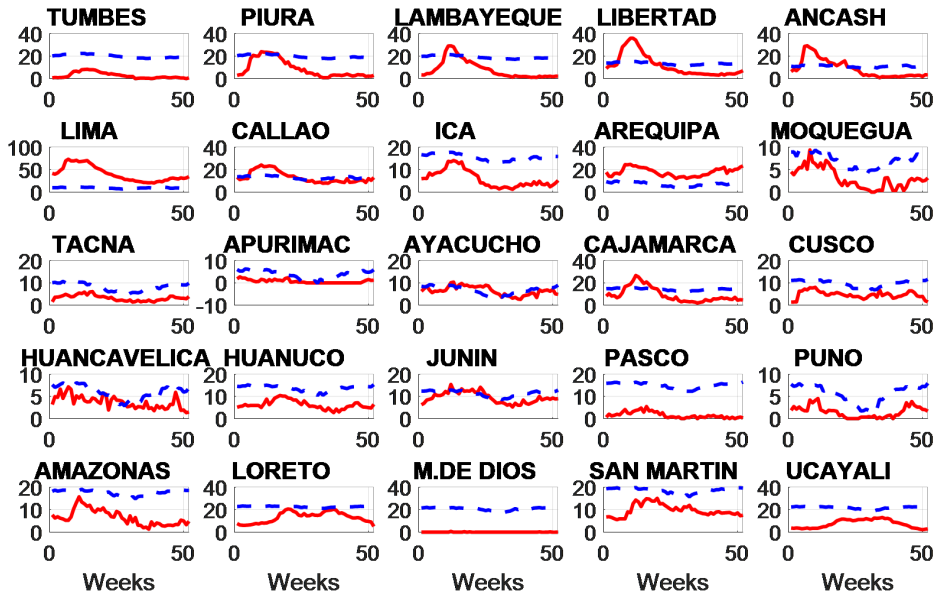


Figure 13: Correlation coefficients for mean case incidence and mean minimum temperature over the first three years of the epidemic. There are moderately strong positive correlations for most coastal and highland departments. Correlations are less consistent for departments in the jungle region.

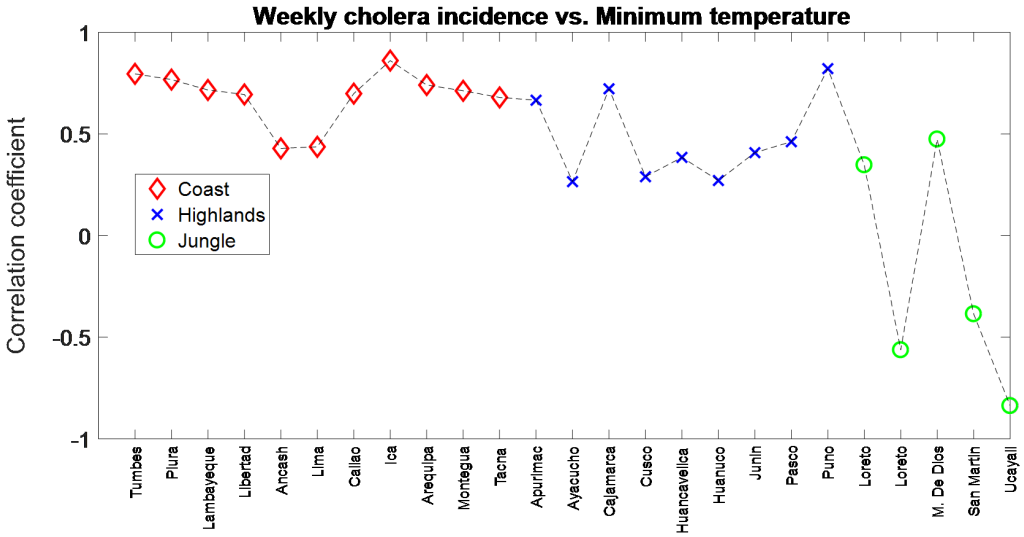


Figure 14: Correlation between average weekly cases and minimum temperatures for two coastal departments (Tumbes and Ica) and two highland departments (Apurimac and Puno).

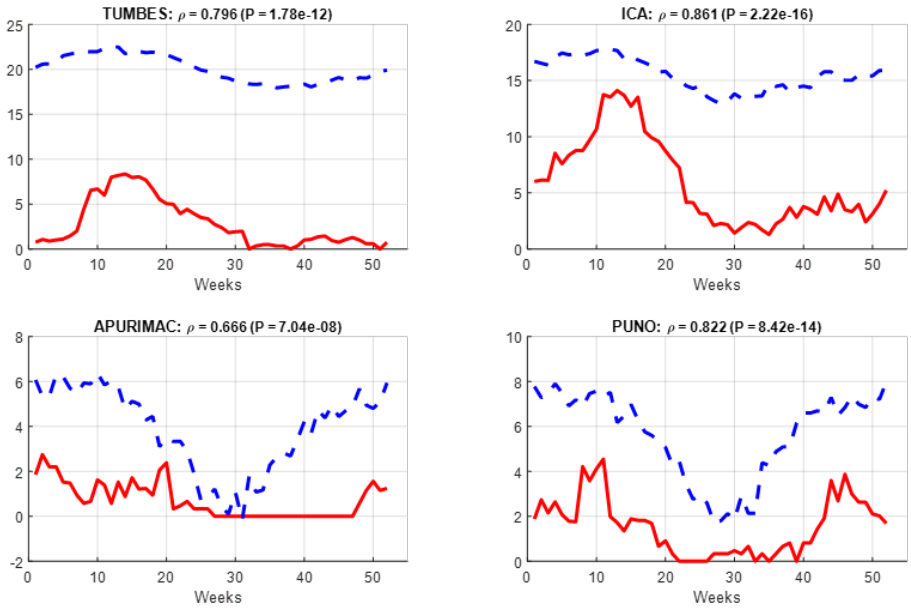


Figure 15: Assessment of model fit by region by evaluating the linear relationship between observed and predicted cases from 1991-1993.

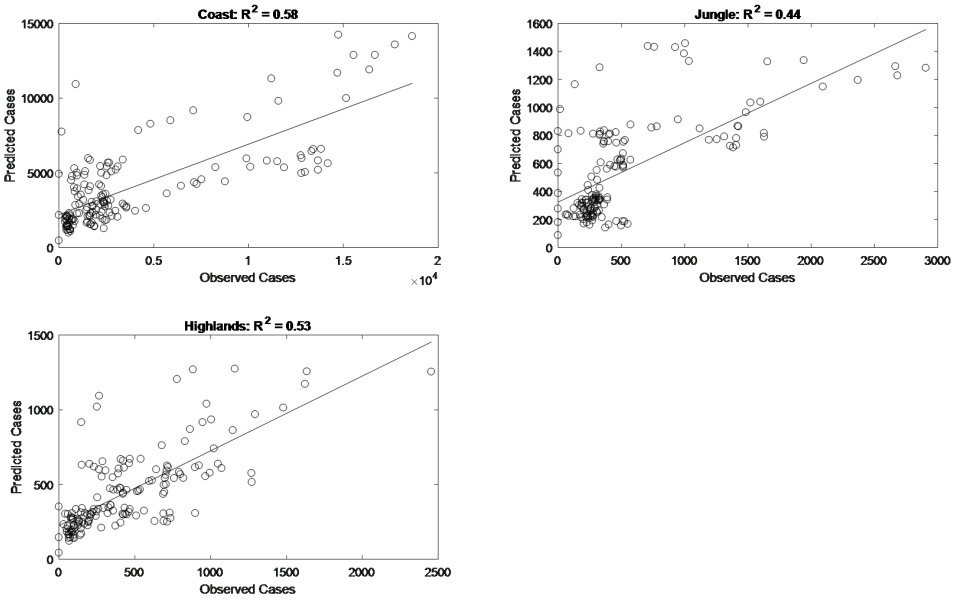


Figure 16: Model fit for 24 departments using approximate Bayesian computation. Blue lines are the model fit using the best estimates for all five parameters, and black lines are the original cholera case counts.

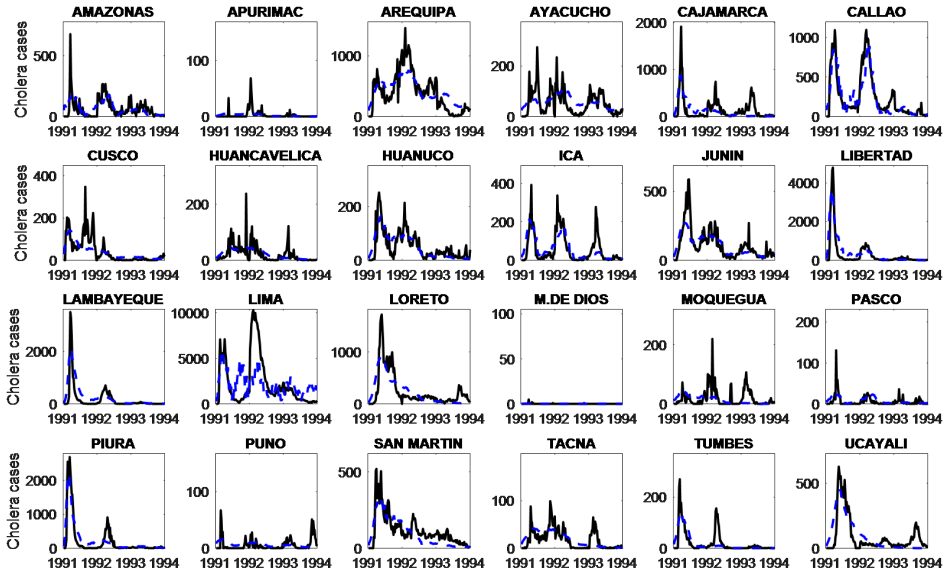


Figure 17: Model fit for the three regions using approximate Bayesian computation. Blue lines are the model fit using the best estimates for all five parameters, and black lines are the original cholera case counts.

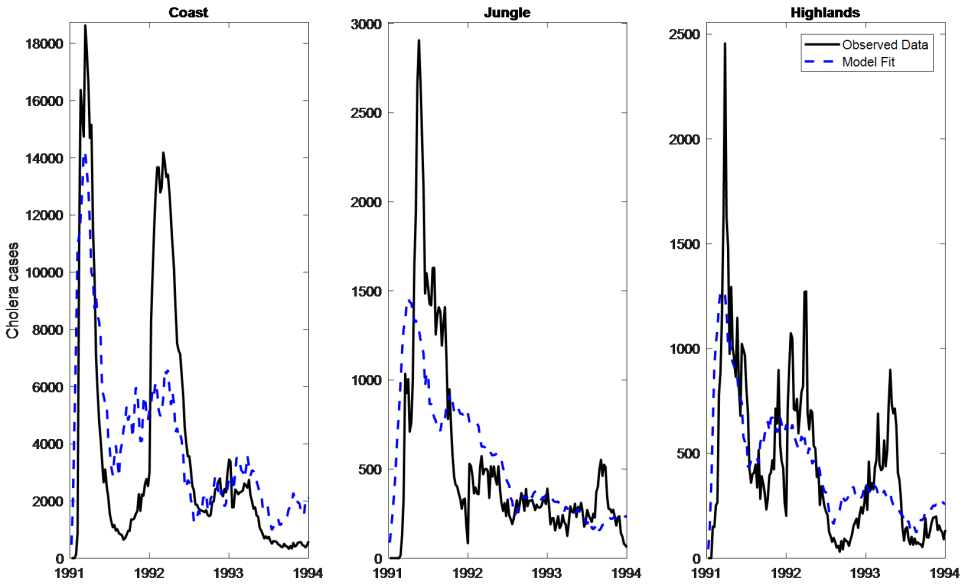


Figure 18: Environmental effective reproductive numbers by department, 1991-1993.

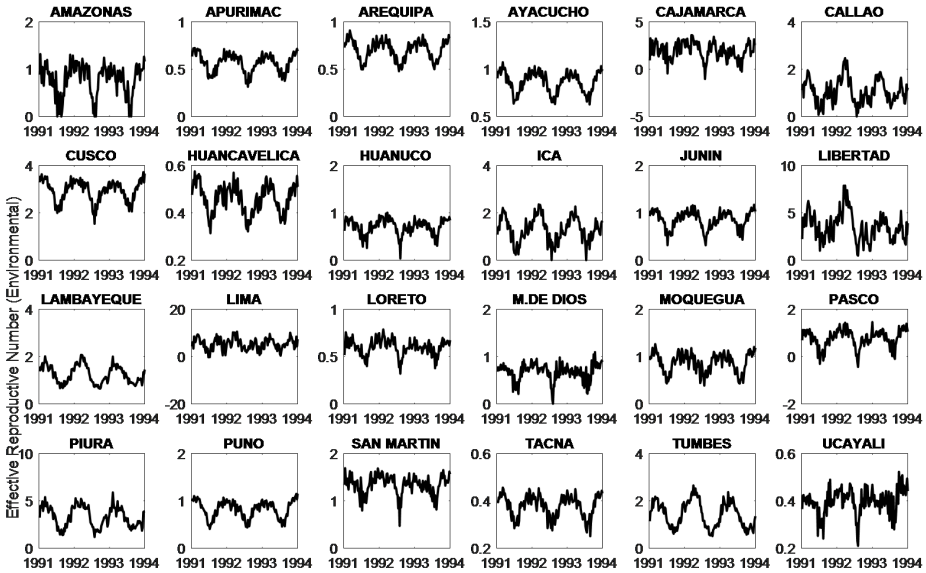
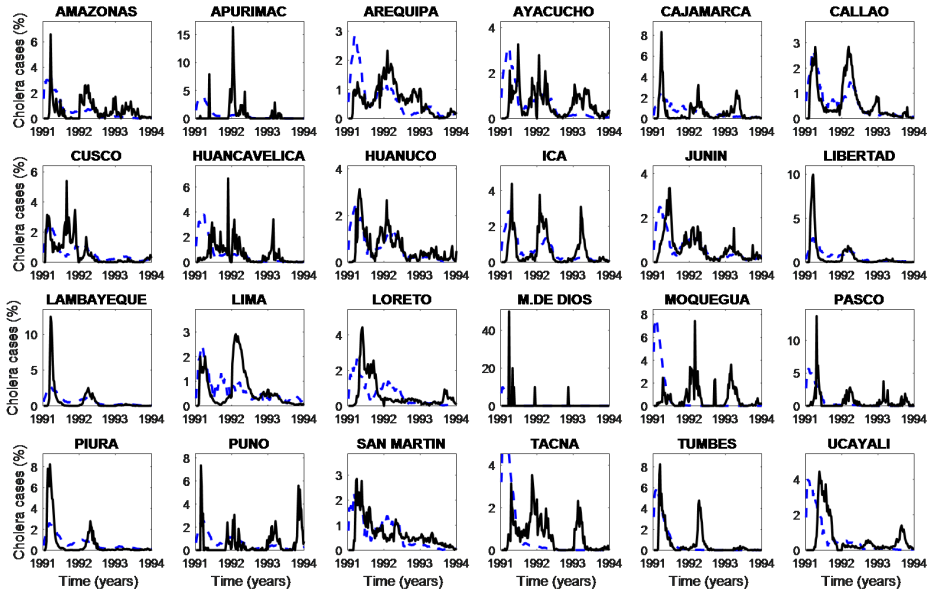


Figure 19: Model fit by department using fixed parameters: $R_h = 0.6$ and $B(0) = 500000$, $\beta_{e(1)} = 0.001$, and $\beta_{e(2)} = 0.00033$. Blue lines are the scaled model fits, and black lines are the scaled case incidence for the 24 departments. By fixing all parameters and allowing only the temperature component to vary, we saw that the model is still able to capture the temporal waves characteristic of cholera transmission.



8 Appendix II

Supplementary Material

Table S1: Department level geographic variables.

	Population	Elevation (meters)	Latitude (°)	Longitude (°)
Amazonas	345101	2334	-6.22	-77.85
Ancash	981760	3052	-9.53	-77.53
Apurimac	394380	3952	-14.17	-72.76
Arequipa	920819	2335	-16.40	-71.54
Ayacucho	524505	2761	-13.16	-74.22
Cajamarca	1281171	2750	-7.16	-78.51
Callao	623464	1	-12.03	-77.13
Cusco	1053244	3399	-13.53	-71.97
Huancavelica	401663	1950	-12.77	-74.98
Huanuco	657717	1880	-9.93	-76.24
Ica	565844	406	-14.07	-75.73
Junin	1079426	4818	-11.48	-74.98
La Libertad	1262368	34	-8.00	-78.50
Lambayeque	924463	3078	-6.43	-79.87
Lima	6331733	1548	-12.04	-77.03
Loreto	708521	220	-4.00	-78.32
Madre de Dios	63917	3932	-11.99	-70.59
Moquegua	127814	3756	-16.80	-70.80
Pasco	243185	4380	-10.50	-75.30
Piura	1392612	29	-4.99	-80.41
Puno	1090838	3830	-15.84	-70.02
San Martin	534172	3080	-7.20	-76.80
Tacna	213166	562	-17.60	-70.20
Tumbes	151889	7	-3.88	-80.59
Ucayali	305823	120	-9.96	-73.19