https://doi.org/10.1016/j.jtbi.2018.04.041

Modeling the 2016-2017 Yemen Cholera Outbreak with the Impact of Limited Medical Resources

Daihai He [a] Xueying Wang [b] Daozhou Gao [c] Jin Wang [d,*]

[a] Department of Applied Mathematics,
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[b] Department of Mathematics,
Washington State University, Pullman, WA 99164, USA

[c] College of Mathematics and Sciences,
Shanghai Normal University, Shanghai 200234, China

[d] Department of Mathematics,
University of Tennessee at Chattanooga, Chattanooga, TN 37403, USA
[*] Email: jin-wang02@utc.edu

Abstract

We present a mathematical model to investigate the transmission dynamics of the 2016-2017 Yemen Cholera Outbreak. Our model describes the interaction between the human hosts and the pathogenic bacteria, under the impact of limited medical resources. We fit our model to Yemen epidemic data published by the World Health Organization, at both the country and regional levels. We find that the Yemen cholera outbreak is shaped by the interplay of environmental, socioeconomic, and climatic factors. Our results suggest that improvement of the public health system and strategic implementation of control measures with respect to time and location are key to future cholera prevention and intervention in Yemen.

1 Introduction

In October 2016, a cholera outbreak started in Yemen. By March 2017, the outbreak was apparently in decline, but it resurged on April 27, 2017 and remained on-going since then. In just two months, cholera rapidly spread to almost every governorate of the country, with reported cases increasing at an average of 5,000 a day. As of November 19, 2017, more than 945,000 suspected cases and over 2,200 disease induced deaths were reported [24], making it the worst cholera outbreak in modern history.

This severe cholera outbreak is a direct consequence of two years of heavy conflict and war, which have been devastating Yemen and pushing the country towards social, economic and institutional collapse. As a result, two thirds of people in Yemen do not have access to clean drinking water and sanitation services are limited, further increasing the risk of catching cholera. Even worse, the national health system has been nearly destroyed by the intense conflict – a recent survey conducted by the World Health Organization (WHO) revealed that only 45% of the health facilities in Yemen remain functional due to shortages in health professionals and medical supplies [23].

On the other hand, the spread of cholera in Yemen exhibits a highly heterogeneous pattern. For example, almost half of the suspected cholera cases have been reported from four governorates: Amanat Al Asimah, Al Hudaydah, Hajjah and Amran, all of which are located on the west coast

of the country. In contrast, for those governorates located on the east part, such as Sayun, Al Mukalla and Al Maharah, the numbers of cholera cases are notably lower (ranging from nearly 0 to slightly above 1,000). Even after scaling by the population size, the attack rates in the west are still significantly higher than those in the east (see Figure 1). Such different levels of disease prevalence and attack rates could be linked to the geographical and climatic distinctions of the country, where the east region is typically dry with an arid climate, and the west region is more humid with a tropical climate (see Figure 2).

Cholera is a waterborne disease caused by the bacterium *Vibrio cholerae*. It is known to spread rapidly in areas with limited access to clean water and sanitation facilities. The transmission of cholera include both indirect (i.e., environment-to-human) and direct (i.e., human-to-human) routes [11]. The indirect transmission pathway typically takes place when people ingest water or food from the environment that is contaminated by the vibrios. The direct transmission may occur when people have close contacts (such as shaking hands or hugging) with infected individuals, or eat food prepared by individuals with dirty hands. A number of studies have demonstrated that both transmission routes play important roles in shaping the complex epidemic and endemic patterns of cholera (see, e.g., [10, 12, 15, 18, 19]).

There are several important questions that remain unanswered regarding the Yemen cholera outbreak. First, why was there a small epidemic (starting in October 2016) prior to the onset of the major epidemic in April 2017? This is often referred to as a "herald wave" in epidemiological study. But what factors trigger this herald wave phenomenon, and how is it connected to the major epidemic wave later on? Second, what is the underlying mechanism that drives the strongly heterogeneous transmission and spread of cholera in different regions of the country? Additionally, what factors lead to such a severe cholera outbreak with long duration? We speculate that the cholera outbreak in Yemen is shaped by the interplay of biological, environmental, socioeconomic, and climatic factors, and answers to these questions demand a systematic study that incorporate all these components.

In the present paper, we propose a new mathematical model to investigate the Yemen cholera outbreak, as a pilot study toward addressing the aforementioned questions. The model represents the interaction between human hosts and the pathogenic bacteria (i.e., the vibrios) and includes both direct and indirect transmission pathways. The model particularly incorporates the impact of limited medical resources on the transmission and spread of cholera in Yemen. We conduct data fitting of our model by using the weekly epidemiological reports published by WHO [25]. Our results contribute to deeper understanding of the transmission dynamics underlying the Yemen cholera outbreak, and provide useful guidelines for the design of control strategies in future.

2 Methods

We propose a mathematical model based on differential equations to investigate the transmission dynamics of the Yemen cholera outbreak. The model involves five compartments: the susceptible individuals (denoted by S), the infectious individuals (denoted by I), the recovered individuals (denoted by R), the concentration of the pathogenic vibrios in the environment (denoted by B), and the availability of medical resources in the country (denoted by M). For convenience of discussion, we normalize the range of M to [0, 1], with the upper limit $M = 1$ (or 100%) representing the maximum strength of the medical resources available in Yemen. Details of the model are provided in Supplemental Materials, S1.1.

We consider both the direct (human-to-human) and the indirect (environment-to-human) transmission routes [11]. The direct and indirect transmission rates are assumed to both depend on the time and the availability of medical resources, in the form of

$$\begin{aligned} \beta_1(M,t) =& \beta(t) \cdot [a_1 + b_1(1-M)], \\ \beta_2(M,t) =& \beta(t) \cdot [a_2 + b_2(1-M)], \end{aligned} \tag{2.1}$$

where a_i and b_i ($i = 1, 2$) are constants, and β is an exponential cubic spline function characterizing the temporal dependence of the transmission rates.

It is well known that the transmission rates play a critical role in determining the output of an epidemic model and, often, the transmission rates vary from place to place. This is especially true for cholera epidemics [10]. In the present study, we use data fitting to evaluate the key parameters a_i, b_i and the function $\beta(t)$, which determine the direct and indirect transmission rates. Other model parameters and their values are discussed in Supplemental Materials, S1.2.

We employ the algorithm of iterated filtering [6, 7] for the data fitting, implemented in a R software package POMP [8]. In particular, when fitting the function $\beta(t)$, we use a number of nodes (controlled by the parameter n_β) uniformly distributed over the entire epidemic period. To quantify the errors of the data fitting and to measure the quality of our model, we calculate the Akaike Information Criterion (AICc), using the small-sample-size corrected version [1,2]. The values of AICc indicate the trade-off between the quality of the fitting and the complexity of the model. A smaller AICc typically suggests a better model. Particularly, in our data fitting, we estimate how many nodes in the cubic spline function $\beta(t)$ may minimize the AICc.

We utilize the outbreak data in the Yemen Situation Reports published weekly by WHO [25]. These data sets contain the weekly reported suspected cases and cumulative cases (as of November 19, 2017) for each governorate as well as the entire country. We conduct data fitting to four time series: country level weekly reported cases, and regional level (the northwest, the southwest, and the east) weekly reported cases. In this study, the three regions (see Figure 1) are defined as

- East: Longitude + Latitude $> 60°$;
- Northwest: Longitude + Latitude $< 60°$ & Latitude $> 14.5°$;
- Southwest: Longitude + Latitude $< 60°$ & Latitude $< 14.5°$.

Such a division roughly corresponds to the three major climatic zones in Yemen [20].

3 Results

Using the method of iterated filtering, we fit our model to the weekly reported outbreak data published by WHO [25] in four scenarios that include the entire country and the three regions. Figure 3 presents our best results of data fitting based on the computed AICc. In particular, we observe that only a small number of nodes are needed to fit the exponential cubic spline function $\beta(t)$, and that fast convergence of the AICc is achieved when n_β increases from 2 to 10, with $n_\beta = 10$ yielding the most accurate results. In all the four scenarios in Figure 3, we observe reasonably good agreement between our simulation result (red curve) and the reported data (black curve with circles). Particularly, our simulation clearly generates the herald wave; i.e., the minor epidemic wave during November 2016 – March 2017 that preceded the major outbreak starting in April/May 2017.

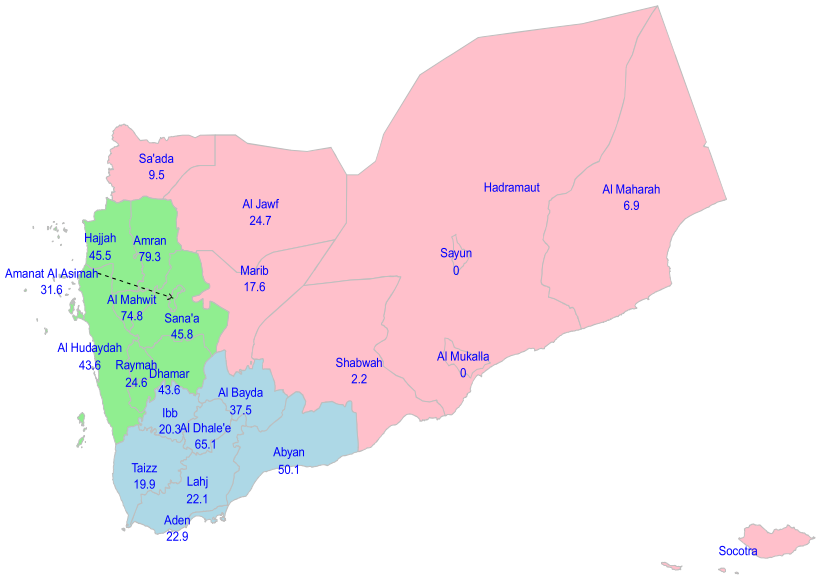

Figure 1: Map of Yemen and the division of three regions: southwest (cyan), northwest (green), and east (pink). The severity of the cholera outbreak is the highest in the northwest, followed by the southwest, and the lowest in the east. Numbers in blue show the attack rates per 1,000 people.

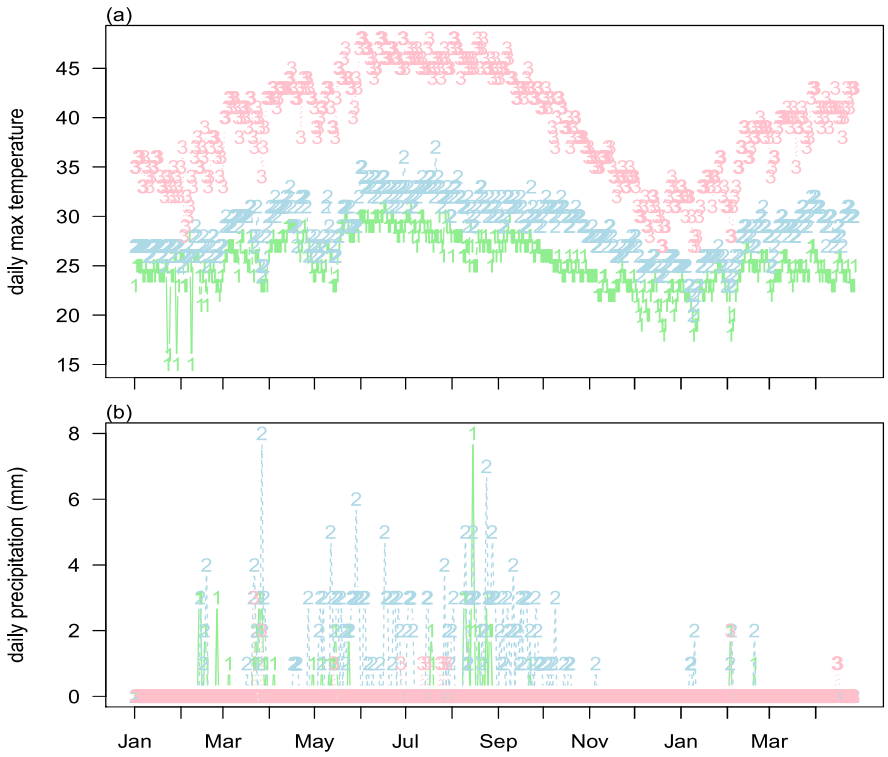

Figure 2: Distribution of temperature and precipitation for the three climatic zones in Yemen. The symbols '1', '2' and '3' refer to the northwest, southwest, and east regions, respectively.

Meanwhile, we have computed the time-dependent reproduction number, $\mathcal{R}_0(t)$, of our mathematical model at a given time t (see Supplemental Materials, S1.3). We find that $\mathcal{R}_0(t)$ has two components: one is the contribution from the human-to-human transmission, denoted by $\mathcal{R}_H(t)$; the other is the contribution from the environment-to-human transmission, denoted by $\mathcal{R}_E(t)$. We treat both $\mathcal{R}_H$ and $\mathcal{R}_E$ as functions of the time variable t, and plot them together in Figure 3.

Figure 3-a shows the result at the country level. At any time, the value of $\mathcal{R}_E$ (represented by the green line with dots) is higher than that of $\mathcal{R}_H$ (represented by the blue dashed line), indicating that the environment-to-human transmission plays a more important role than the human-to-human transmission, on the scale of the entire country. Both $\mathcal{R}_E$ and $\mathcal{R}_H$ attain their peak values between late April and early May, 2017, which coincides with the resurgence of the outbreak (the "major" wave) starting on April 27. They attain their second highest values between October and November, 2016, which corresponds to the onset of the outbreak (the herald wave). Their third highest values occur in September 2017, where the reported data show a moderate increase of infection cases, after a period (about 3 months) of steady decrease. At any time t, we see that $\mathcal{R}_0 = \mathcal{R}_E + \mathcal{R}_H > 1$.

The results for the southwest region (Figure 3-b) and the northwest region (Figure 3-c) show a similar pattern to those in the country level. For the east region (Figure 3-d), however, we observe that the value of $\mathcal{R}_H$ is always higher than that of $\mathcal{R}_E$, indicating the dominance of the human-to-human transmission in this particular region.

Additionally, it is easy to observe that the northwest region has the highest $\mathcal{R}_0$ (i.e., $\mathcal{R}_E$ and $\mathcal{R}_H$ combined) at any time, among the three regions, followed by the southwest region, and the east region has the lowest $\mathcal{R}_0$. This observation is consistent with the fact that the northwest and east regions exhibit the highest and lowest degrees of outbreak severity, respectively.

4 Discussion

The transmission and spread of cholera involve complex biological and socioeconomic processes, and the underlying mechanisms for a cholera outbreak vary from place to place. In this study, we have attempted to use mathematical modeling and data fitting to investigate the Yemen cholera outbreak, with an aim to better understand the patterns and driving forces for this extremely severe and long lasting cholera epidemic.

The unusually high prevalence and severity of this cholera outbreak are certainly linked to the collapsed health system and limited medical resources in Yemen. In our work, we have incorporated the strength of the medical resources as an essential variable in the model, on which the transmission rates explicitly depend. Another important feature of our model is that we have introduced a time-dependent function $\beta(t)$ which can characterize the transmission rates at different times due to the change of environmental, socioeconomic and climatic conditions. In particular, the function $\beta(t)$ takes different values at the three different regions of the country, representing the impact of spatial and geographical heterogeneity on cholera transmission. These features enable us to fit our model to realistic data and obtain meaningful fitting results.

We have employed time-dependent reproduction numbers, $\mathcal{R}_E$ and $\mathcal{R}_H$, as a means to quantify the disease risks throughout the epidemic period. We find that the values of $\mathcal{R}_E$ and $\mathcal{R}_H$ vary significantly among the three different regions, a clear indication that the different levels of epidemic severity are linked to the geographical and climatic heterogeneity. For example, the east region is characterized by the arid (or, desert) climate with very low precipitation throughout the year [20]. Such a dry environment is "hostile" to the growth and survival of *Vibrio cholerae*. It is

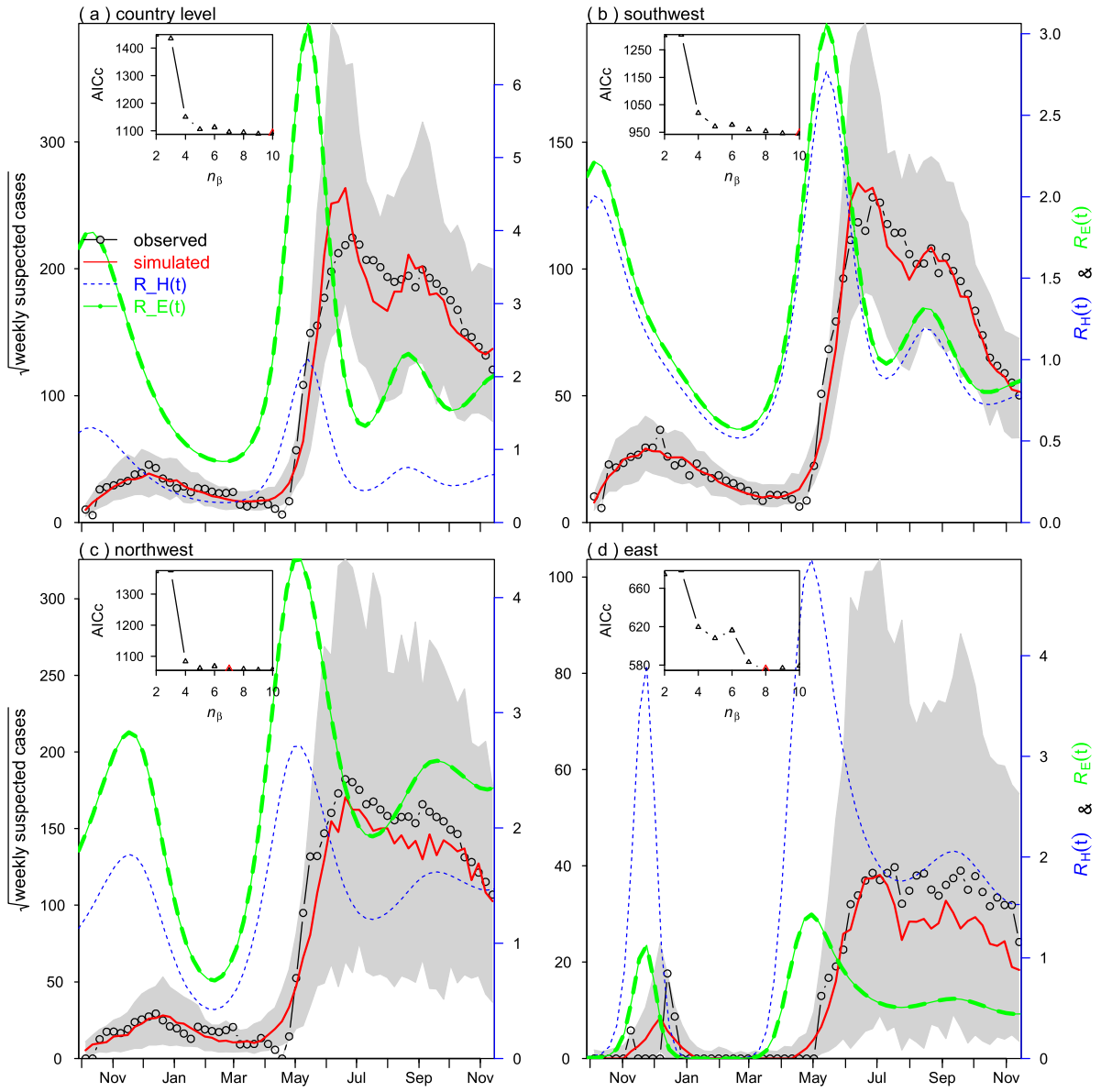

Figure 3: Data fitting results in four scenarios: (a) the entire country; (b) the southwest region; (c) the northwest region; (d) the east region. Black line with circles denotes reported cases, red line denotes model simulation median, blue dashed line and green line with dots show the time evolution of $\mathcal{R}_H$ and $\mathcal{R}_E$, respectively, and shaded region represents 95% bound of 1,000 model simulations.

thus not surprising to see that the environment-to-human transmission route plays a minor role in this region, whereas the human-to-human transmission pathway accounts for the majority of the disease prevalence. In contrast, the southwest and northwest regions are much more humid, with a tropical climate, which tends to promote the transmission and spread of the vibrios, particularly in April/May when the precipitation optimizes the growth of the vibrios. For example, the capital Sana'a, a city located in the northwest region and heavily plagued by the cholera outbreak, has an average rainfall of 1 mm in each month of October, November, December and January, 9 mm in February, 7 mm in March, and 35 mm and 13 mm in April and May, respectively [21]. Thus, the

months of April and May mark a period with a sharp increase in precipitation, coinciding with the onset of the major epidemic that started in Sana'a on April 27, 2017 and the fact that the sewer system in the city stopped working just a few days before.

Through data fitting, we obtain simulation results that closely represent the first, minor epidemic (the herald wave) and the second, severe epidemic (the major wave). The herald wave from November 2016 to March 2017 is accompanied by relatively high values of $\mathcal{R}_E$ and $\mathcal{R}_H$ in the beginning. As both $\mathcal{R}_E$ and $\mathcal{R}_H$ decrease, the wave is in decline in the following months. However, a sudden increase of $\mathcal{R}_E$ and $\mathcal{R}_H$ to their peak values in late April triggers the major epidemic wave. Such a rapid change of $\mathcal{R}_E$ and $\mathcal{R}_H$ is possibly attributed to the change of both environmental and socioeconomic conditions. For most parts of Yemen, on top of the higher precipitations in April/May, the temperature in April/May could reach mid 30s Celsius which is optimal for the growth and reproduction of *Vibrio cholerae* [3, 14]. Meanwhile, as the war continues, more health facilities and medical resources are destroyed, and public infrastructures get damaged (particularly, the dysfunction of Sana'a's sewer system in April 2017), further increasing the risk of cholera transmission and rendering people in Yemen more vulnerable to the infection.

Our study indicates that several major factors collectively shape the epidemics of the Yemen cholera outbreak. First, the collapsed health system and limited medical resources directly impact the transmission and spread of the disease, and transmission rates increase when the strength of the medical resources decays. Second, the evolution of both the direct and indirect transmission pathways with time leads to a peak level of transmission rates in April/May 2017 that triggers the major epidemic. Third, the geographical and climatic distinctions among the west and east regions account for the heterogeneous epidemic patterns throughout the country.

These results have important implications to future cholera prevention/intervention in Yemen. Improvement of public health infrastructure and recovery of medical facilities are of fundamental importance. In addition, control efforts should be implemented strategically with respect to time and location: Stronger control measures should be planned for the time periods (such as April/May) when the temperature and precipitation optimize the growth of the pathogen, and the west part of the country (particularly the northwest region) demands a higher level of control than the east part.

Finally, there are several limitations in our study. First, our mathematical epidemic model is a non-autonomous system, for which there is no general theory to define or compute the basic reproduction number (note that the standard next-generation matrix framework for autonomous systems [4,16] and some extended work for periodic systems [12,17] do not apply to our model). We have used a simple way to evaluate $\mathcal{R}_H$ and $\mathcal{R}_E$ by freezing the system at each time t and treating it as an autonomous system at t. Although this approach lacks mathematical rigor, it provides an intuitive means to measure the risks of infection (due to the direct and indirect transmissions, respectively) that change with time. Second, in our data fitting, we set the recruitment rate, Λ, for the medical resources as 0. This simplifies the task of data fitting, but may not realistically represent the situation in Yemen. Additionally, in reality, several other model parameters such as the bacterial shedding rate and the disease recovery rate, may also take complex forms and may vary with time and space, and such complications are not explored in the current study.

Acknowledgments

DH was supported by the Early Career Scheme from Hong Kong Research Grants Council (PolyU 251001/14M). XW was supported in part by a grant from the Simons foundation (No. 317407). DG was supported by NSFC (No. 11601336), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (TP2015050), and Shanghai Gaofeng Project for University Academic Development Program. JW was partially supported by the National Science Foundation under Grant Nos. 1412826 and 1557739. The authors are grateful to the two anonymous referees for their helpful comments.

References

[1] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19(6): 716-723, 1974.

[2] H. Akaike, Prediction and entropy, in *A Celebration of Statistics*, A.C. Atkinson and S.E. Fienberg (Ed.), Springer, pp. 1-24, 1985.

[3] G.C. de Magny and R.R. Colwell, Cholera and climate: A demonstrated relationship, *Transactions of the American Clinical and Climatological Association*, 120: 119-128, 2009.

[4] O. Diekmann, J.A.P. Heesterbeek, and A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous population, *Journal of Mathematical Biology*, 28: 365-382, 1990.

[5] D.M. Hartley, J.G. Morris, and D. L. Smith, Hyperinfectivity: A critical element in the ability of *V. cholerae* to cause epidemics? *PLoS Medicine*, 3: 0063-0069, 2006.

[6] E.L. Ionides, A. Bhadra, Y. Atchade, and A.A. King, Iterated filtering, *Annals of Statistics*, 39(3): 1776-1802, 2011.

[7] A.A. King, E.L. Ionides, M. Pascual, and M.J. Bouma, Inapparent infections and cholera dynamics, *Nature*, 454: 877-880, 2008.

[8] A.A. King, D. Nguyen, and E.L. Ionides, Statistical inference for partially observed Markov processes via the R package pomp, *Journal of Statistical Software*, 69(12): 1-43, 2016.

[9] J.P. LaSalle, The stability of dynamical systems, *Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, 1976.

[10] Z. Mukandavire, S. Liao, J. Wang, H. Gaff, D.L. Smith, and J.G. Morris, Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in Zimbabwe, *Proceedings of the National Academy of Sciences*, 108: 8767-8772, 2011.

[11] E.J. Nelson, J.B. Harris, J.G. Morris, S.B. Calderwood, and A. Camilli, Cholera transmission: The host, pathogen and bacteriophage dynamics, *Nature Reviews: Microbiology*, 7: 693-702, 2009.

[12] D. Posny and J. Wang, Modeling cholera in periodic environments, *Journal of Biological Dynamics*, 8(1): 1-19, 2014.

[13] Z. Shuai and P. van den Driessche, Global stability of infectious disease models using Lyapunov functions, *SIAM Journal on Applied Mathematics*, 73: 1513-1532, 2013.

[14] F.L. Singleton, R. Attwell, S. Jangi, and R.R. Colwell, Effects of temperature and salinity on *Vibrio cholerae* growth, *Applied and Environmental Microbiology*, 44: 1047-1058, 1982.

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

[16] P. van den Driessche and J. Watmough, Reproduction number and subthreshold endemic equilibria for compartment models of disease transmission, *Mathematical Biosciences*, 180: 29-48, 2002.

[17] W. Wang and X.-Q. Zhao, Threshold dynamics for compartmental epidemic models in periodic environments, *Journal of Dynamics and Differential Equations*, 20: 699-717, 2008.

[18] X. Wang, D. Posny and J. Wang, A reaction-convection-diffusion model for cholera spatial dynamics, *Discrete and Continuous Dynamical Systems B*, 21: 2785-2809, 2016.

[19] C. Yang, X. Wang, D. Gao, and J. Wang, Impact of awareness programs on cholera dynamics: Two modeling approaches, *Bulletin of Mathematical Biology*, 79: 2109-2131, 2017.

[20] Climate and Average Weather in Yemen: http://weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine-in-Yemen

[21] World Climate Guide – Yemen: http://www.climatestotravel.com/climate/yemen

[22] WHO Global health observatory data repository: Life expectancy, 2013. Available from: http://apps.who.int/gho/data/view.main.680?lang=en

[23] WHO survey on Yemen's health system: http://www.emro.who.int/media/news/survey-reveals-extent-of-damage-to-yemens-health-system.html

[24] WHO Weekly Epidemiology Bulletin, 13-19 November 2017.

[25] WHO Yemen cholera situation reports: http://www.emro.who.int/yem/yemeninfocus/situation-reports.html

[26] Wikipedia page for Yemen: http://en.wikipedia.org/wiki/Yemen

Supplemental Materials

S1.1 Mathematical model

Let S, I and R be the numbers of susceptible, infectious and recovered individuals, respectively, and B the concentration of the vibrios in the environment. Let also M represent the availability of related medical resources, which could include health centers, sanitation systems, medicines, etc. We propose the following differential equation model for the Yemen cholera outbreak:

$$\begin{aligned}
\frac{dS}{dt} &= \mu N - \beta_1(M,t)S\frac{I}{N} - \beta_2(M,t)S\frac{B}{B+K} - \mu S,\\
\frac{dI}{dt} &= \beta_1(M,t)S\frac{I}{N} + \beta_2(M,t)S\frac{B}{B+K} - \gamma(M)I - \mu I,\\
\frac{dR}{dt} &= \gamma(M)I - \mu R,\\
\frac{dB}{dt} &= \xi I - \delta B,\\
\frac{dM}{dt} &= \Lambda - \alpha M + \phi\frac{I}{N}.
\end{aligned} \tag{S1}$$

Here the constant $N = S(t) + I(t) + R(t)$ is the total population size of human hosts, the parameter μ is the natural birth and death rate for the human hosts, K is the half saturation concentration of the vibrios, ξ is the rate of human contribution (typically through shedding) to the environmental vibrios, and δ is the natural removal rate of vibrios from the environment. The availability of public health resources has an influx rate Λ, and a decay rate α due to the ongoing conflict and violence. Meanwhile, it is stimulated by the disease prevalence at a rate ϕ due to international assistance from WHO, UNICEF, etc. The parameters β_1 and β_2 are the direct (or, human-to-human) and indirect (or, environment-to-human) transmission rates, respectively, and γ is the rate of recovery from the infection. We assume that β_1, β_2, and γ are all explicitly depending on M, and that β_1 and β_2 both increase when M decays, reflecting the impact of the collapsing public health system on cholera transmission and spread. Specifically, we make the following assumptions:

(H1) $\beta_1(M,t)$, $\beta_2(M,t)$ and $\gamma(M)$ are continuously differentiable;

(H2) $\frac{\partial}{\partial M}\beta_1(M,t) \le 0$, $\frac{\partial}{\partial M}\beta_2(M,t) \le 0$, and $\gamma'(M) \ge 0$.

The parameters $\beta_1(M,t)$ and $\beta_2(M,t)$ are defined in equation (2.1). Meanwhile, we assume that $\gamma(M) = a_3 + b_3 M$ with $a_3 > 0$ and $b_3 > 0$, reflecting the increase of the recovery rate when the medical resources are improved.

S1.2 Parameter values

Values of the parameters in model (S1) are provided in Table 1. The entire population in Yemen is about $27,540,666$ as of 2016 [26]. Meanwhile, the populations aggregated in the southwest, northwest and east regions are about $9,949,014$, $14,375,918$ and $2,594,993$, respectively. The natural human birth rate and death rate take the same value, μ, in our model. The average life expectancy in Yemen is about 64 years [22], and so we may estimate μ by $\mu \approx (64 \text{ year})^{-1} \approx 0.0156 \text{ year}^{-1}$. *Vibrio cholerae* can typically survive in the aquatic environment for about 30

days [11], thus we calculate the natural death (or, removal) rate of the bacteria by $\delta \approx (30 \text{ day})^{-1} \approx 12.167 \text{ year}^{-1}$. Based on the recent WHO survey on Yemen's health system [23], only 45% of the medical facilities in Yemen remain fully functional after two years of conflicts. Then we may roughly estimate the decay rate of the medical resources by $\alpha \approx (-0.5) * \ln(0.45) \approx 0.399 \text{ year}^{-1}$. An individual infected with cholera typically recovers in $5 - 20$ days [5]. Hence, we compute the minimal recovery rate by $a_3 \approx (20 \text{ day})^{-1} \approx 18.25 \text{ year}^{-1}$, and the maximal recovery rate by $a_3 + b_3 \approx (5 \text{ day})^{-1} \approx 73.0 \text{ year}^{-1}$. Note that M is normalized to the range of [0, 1] in our data fitting. Consequently, we obtain $b_3 \approx 54.75 \text{ year}^{-1}$. According to [7], asymptomatic infections could account for $75 - 99\%$ of the total cholera cases, thus we assume a 20% reporting ratio of the cholera data in Yemen. Those parameters related to the disease transmission rates, a_i, b_i $(i = 1, 2)$ and $\beta(t)$, are determined through data fitting. The transmission rates β_1 and β_2 thus take different values at the country level and in the three different regions. Additional parameter values are either directly found through literature or assumed.

Table 1: Definitions and values of model parameters

Parameter	Definition	Value	Reference
N	Total human population	$27,540,666$ persons	[26]
μ	Natural human birth/death rate	$0.0156\,\text{year}^{-1}$	[22]
K	Half saturation rate	$10^6\,\text{cells} \cdot \text{ml}^{-1}$	[5]
ξ	Bacterial shedding rate	$10\,\text{cells} \cdot \text{ml}^{-1}\text{person}^{-1}$	[5]
δ	Bacterial removal rate	$12.167\,\text{year}^{-1}$	[11]
Λ	Influx rate of medical resources	$0\,\text{year}^{-1}$	assumed
α	Decay rate of medical resources	$0.399\,\text{year}^{-1}$	[23]
θ	Reporting ratio of cholera cases	20%	[7]
ϕ	Stimulation rate of medical resources	$0.108\,\text{year}^{-1}$	data
a_1	Base component of direct transmission rate	1.0	data
b_1	Medically impacted direct transmission rate	10.0	data
a_2	Base component of indirect transmission rate	1.0	data
b_2	Medically impacted indirect transmission rate	10.0	data
$\beta(t)$	Time-dependent component of transmission rates	vary in time	data
n_β	Number of nodes for $\beta(t)$	vary by simulation	[6]
a_3	Minimal value of recovery rate	$18.25\,\text{year}^{-1}$	[5]
b_3	Medically impacted component of recovery rate	$54.75\,\text{year}^{-1}$	[5]

S1.3 Time-dependent reproduction number

The concept of the basic reproduction number, which measures the expected number of secondary infection produced by one primary case, has been commonly used in epidemic models based on autonomous dynamical systems. However, due to the explicit time dependence of the two transmission rates, our proposed model (S1) is a non-autonomous system for which a general procedure to compute and analyze the basic reproduction number is not available. For simplicity, we treat our model as an autonomous system at a given time t, with the transmission rates β_1 and β_2 frozen at t. Using the next-generation matrix technique [16], we define the time-dependent reproduction number $\mathcal{R}_0(t)$ for our model as the spectral radius of the next generation matrix at time t, which

is symbolically computed as

$$\mathcal{R}_0(t) = \frac{\beta_1(M_0,t)}{\gamma(M_0)+\mu} + \frac{\beta_2(M_0,t)\xi N}{[\gamma(M_0)+\mu]\delta K} := \mathcal{R}_H(t) + \mathcal{R}_E(t)\,, \tag{S2}$$

where $M_0 = \Lambda/\alpha$. The quantities $\mathcal{R}_H(t)$ and $\mathcal{R}_E(t)$ measure the contributions from the human-to-human and the environment-to-human transmissions routes, respectively, to the overall infection risk at time t.

S1.4 Equilibrium analysis

To gain more analytical insight into our model, we conduct a mathematical analysis for the simplified system where we assume that the two transmission rates are independent of time: $\beta_1(M,t) = \beta_1(M)$ and $\beta_2(M,t) = \beta_2(M)$. Consequently, the time-dependent reproduction number defined in equation (S2) becomes a constant, $\mathcal{R}_0$.

Let

$$\Omega = \{(S,I,R,B,M) \in \mathbb{R}^5_+ : S+I+R = N, 0 \le B \le B_{max}, 0 \le M \le M_{max}\}, \tag{S3}$$

with $B_{max} = \xi N/\delta$ and $M_{max} = (\Lambda+\phi)/\alpha$. It is easy to see that Ω is positively invariant; that is, any solution of system (S1) that is initially in Ω will remain in Ω in forward time. Hence, any biologically feasible solution of (S1) is nonnegative and bounded.

It is obvious that the system (S1) admits a unique disease-free equilibrium (DFE), given by $(S,I,R,B,M) = (N,0,0,0,\Lambda/\alpha) := E_0$. We proceed to analyze the non-trivial equilibrium solutions of model (S1). Let $E_e := (S_e, I_e, R_e, B_e, M_e)$ denote an equilibrium solution of model (S1) with $E_e \neq E_0$. Clearly, E_e satisfies

$$\begin{aligned}
&\mu N - \beta_1(M)S\frac{I}{N} - \beta_2(M)S\frac{B}{B+K} - \mu S = 0,\\
&\beta_1(M)S\frac{I}{N} + \beta_2(M)S\frac{B}{B+K} - \gamma(M)I - \mu I = 0,\\
&\gamma(M)I - \mu R = 0,\\
&\xi I - \delta B = 0,\\
&\Lambda - \alpha M + \phi\frac{I}{N} = 0.
\end{aligned} \tag{S4}$$

These imply

$$\begin{aligned}
S &= \frac{(\gamma(M)+\mu)I}{\beta_1(M)\frac{I}{N} + \beta_2(M)\frac{B}{B+K}},\\
R &= \frac{\gamma(M)I}{\mu},\\
B &= \frac{\xi I}{\delta},\\
M &= \frac{\Lambda + \phi I/N}{\alpha} := \theta(I).
\end{aligned} \tag{S5}$$

Solving (S5) yields

$$S = U(I) := \frac{\gamma(\theta(I))+\mu}{\frac{\beta_1(\theta(I))}{N} + \beta_2(\theta(I))\frac{\xi}{\xi I + K\delta}}. \tag{S6}$$

One can verify by direct calculation that $U(I)$ is an increasing function. On the other hand, $S = N - I - R = N - (1 + \frac{\gamma(\theta(I))}{\mu})I := V(I)$ is strictly decreasing, and hence we know a nontrivial equilibrium solution can be determined by the intersection of $U(I)$ and $V(I)$ in $\mathring{\mathbb{R}}^2_+$, where $\mathring{\mathbb{R}}^2_+$ denotes the interior of $\mathbb{R}^2_+$. By virtue of $U(0) = N/\mathcal{R}_0$ and $V(0) = N$, we find that (1) if $\mathcal{R}_0 \leq 1$, the two curves $U(I)$ and $V(I)$ do not intersect in $\mathring{\mathbb{R}}^2_+$; (2) if $\mathcal{R}_0 > 1$, they have a unique intersection in $\mathring{\mathbb{R}}^2_+$. This implies that there exist at most two biologically feasible equilibria for model (S1) which depends on the value of $\mathcal{R}_0$; namely, *the DFE is the only equilibrium when* $\mathcal{R}_0 \leq 1$*, whereas there are two equilibria, the DFE and the endemic equilibrium (EE), when* $\mathcal{R}_0 > 1$.

The theorem below summarizes the result on disease extinction and persistence by establishing $\mathcal{R}_0$ as a sharp threshold for the disease dynamics.

Theorem S1.1. *The following conclusions hold for model* (S1)*:*

1. *If* $\mathcal{R}_0 \leq 1$*, then the DFE is globally asymptotically stable in* Ω*;*
2. *If* $\mathcal{R}_0 > 1$*, then the DFE is unstable and the system is uniformly persistent in* $\mathring{\Omega}$*; that is,* $\liminf_{t\to\infty} I(t) \geq c$ *and* $\liminf_{t\to\infty} B(t) \geq c$ *for some* $c > 0$*, where* $\mathring{\Omega}$ *denotes the interior of* Ω*.*

Proof. Take $x = (I, B)^T$ and $y = (S, R, M)^T$. Denote

$$f(x, y) = (F - V)x - \begin{pmatrix} \beta_1(M)S\frac{I}{N} + \beta_2(M)S\frac{B}{B+K} - \gamma(M)I - \mu I \\ \xi I - \delta B \end{pmatrix} \tag{S7}$$

where F and V are defined as

$$F = \begin{bmatrix} \beta_1(M_0) & \beta_2(M_0)\dfrac{N}{K} \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} \gamma(M_0) + \mu & 0 \\ -\xi & \delta \end{bmatrix}. \tag{S8}$$

One can verify that $f(x, y) \geq 0$ in Ω, $F \geq 0$, $V^{-1} \geq 0$ and $V^{-1}F$ is irreducible, where $f \geq 0$ in Ω is obtained using assumptions (H1)-(H2). By Theorem 2.2 of [13], all the assertions follow when $\mathcal{R}_0 \neq 1$.

It remains to prove the case when $\mathcal{R}_0 = 1$. Let $w = (\beta_1(M_0)N,\ \beta_2(M_0)N/K)$. It is easy to see from direct calculation that w is a left eigenvector of $V^{-1}F$ associated with the eigenvalue $\mathcal{R}_0 = \rho(FV^{-1}) = \rho(V^{-1}F)$; namely, $wV^{-1}F = w\mathcal{R}_0$. Inspired by [13], we consider the Lyapunov function $\mathcal{L} = w\,V^{-1}x$. Differentiating $\mathcal{L}$ with respect to the solution of model (S1) leads to

$$\mathcal{L}' = w\,V^{-1}\frac{dx}{dt} \leq w\,V^{-1}(F - V)x = w(\mathcal{R}_0 - 1)x = 0, \tag{S9}$$

where the inequality follows from $f \geq 0$ in Ω. Thus, $\mathcal{L}' = 0$, and this leads to $\beta_1(M)SI/N = \beta_1(M_0)I$, $\beta_2(M)SB/(B + K) = \beta_2(M_0)NB/K$. It implies that (C1) $I = B = 0$, or (C2) $S = N$, $\beta_1(M) = \beta_1(M_0)$, $B = 0$. When (C1) holds, equation (S1) yields $S = N$, $M = M_0$ and $R = 0$. Similarly, when (C2) is valid, we obtain $I = 0$, which can be reduced to the case (C1). Therefore, the largest invariant set such that $\mathcal{L}' = 0$ is $\{E_0\}$. It follows from LaSalle's Invariance Principle [9] that the DFE (i.e., E_0) is globally asymptotically stable in Ω. ∎

Theorem S1.1 essentially states that the disease dies out when the reproduction number is lower than or equal to unity, and the disease persists when the reproduction number is higher than unity.