

RESEARCH

Open Access



# Transmission dynamics of cholera in Yemen, 2017: a real time forecasting

Hiroshi Nishiura<sup>1,2\*</sup> , Shinya Tsuzuki<sup>1,2</sup>, Baoyin Yuan<sup>1,2</sup>, Takayuki Yamaguchi<sup>1,2</sup> and Yusuke Asai<sup>1,2</sup>

\* Correspondence:

nishiurah@med.hokudai.ac.jp

<sup>1</sup>Graduate School of Medicine,  
Hokkaido University, Kita 15 Jo Nishi,  
7 Chome, Kita-ku, Sapporo  
060-8638, Japan

<sup>2</sup>CREST, Japan Science and  
Technology Agency, 4-1-8, Honcho,  
Kawaguchi-shi, Saitama 332-0012,  
Japan

## Abstract

**Background:** A large epidemic of cholera, caused by *Vibrio cholerae*, serotype Ogawa, has been ongoing in Yemen, 2017. To improve the situation awareness, the present study aimed to forecast the cholera epidemic, explicitly addressing the reporting delay and ascertainment bias.

**Methods:** Using weekly incidence of suspected cases, updated as a revised epidemic curve every week, the reporting delay was explicitly incorporated into the estimation model. Using the weekly case fatality risk as calculated by the World Health Organization, ascertainment bias was adjusted, enabling us to parameterize the family of logistic curves (i.e., logistic and generalized logistic models) for describing the unbiased incidence in 2017.

**Results:** The cumulative incidence at the end of the epidemic, was estimated at 790,778 (95% CI: 700,495, 914,442) cases and 767,029 (95% CI: 690,877, 871,671) cases, respectively, by using logistic and generalized logistic models. It was also estimated that we have just passed through the epidemic peak by week 26, 2017. From week 27 onwards, the weekly incidence was predicted to decrease.

**Conclusions:** Cholera epidemic in Yemen, 2017 was predicted to soon start to decrease. If the weekly incidence is reported in the up-to-the-minute manner and updated in later weeks, not a single data point but the entire epidemic curve must be precisely updated.

**Keywords:** *Vibrio cholerae*, Outbreak, Epidemiology, Prediction, Backcalculation

## Background

Cholera is a bacterial disease caused by infection of small intestine by *Vibrio cholerae*, characterized by a variety of diarrhea, abdominal cramp and dehydration. Most common route of infection is through contaminated water and foods, and due to the environmental nature of transmission, the control is complicated in tropical environment where clean water is not easily accessible. From 2016, a large epidemic of cholera has been seen in Yemen, and stool samples have tested positive for *V. cholerae*, serotype Ogawa. The incidence once declined in early 2017, but a bigger epidemic was triggered from early-mid April 2017 with unprecedented size of cases [1]. Essential resources including vaccines, fluids and antibiotics have been allocated.

Mathematical modeling studies have contributed to better understanding of the cholera transmission dynamics [2–7]. Many models during the 2010 Haitian cholera outbreak explicitly accounted for the presence of asymptomatic individuals and also

the transmission through water environment, because these features, especially the former have been the issue of unrecognized part of the outbreak [8]. Models have been frequently utilized for theoretically optimizing resource allocation of antibiotics and oral vaccines [2–8] in collaboration with public health experts [9].

Such modeling effort has been extended to real-time analysis and explicit assessment of predictive ability during the outbreak in Haiti [10, 11]. However, in the context of Yemen epidemic 2017, a modeling effort for both now-casting and forecasting, elevating situation awareness of both experts and the public, has yet to be made using a parsimonious modeling approach, possibly with real time updates. The present study aims to forecast the cholera epidemic in Yemen, 2017, explicitly addressing the reporting delay and ascertainment bias.

## Methods

### Epidemiological data

The present study analyzed the 2017 epidemic that we assumed to have started from 16 April 2017 (the first date of week 16) which coincided with a new increase in the reported incidence in 2017. Situation report of the cholera outbreak, as represented by Weekly Update and Epidemiology Bulletin, has been managed by the World Health Organization (WHO) Regional Office for the Eastern Mediterranean Regional Office, and the present study used weekly counts of suspected cases and deaths as reported in the Epidemiology Bulletin [12]. Suspected case of cholera adhered to the WHO's classical case definition. That is, a case of cholera was suspected when a patient aged 5 years or more develops acute watery diarrhea, with or without vomiting [13]. The so-called case fatality risk (CFR), which was calculated in each week as the ratio of weekly count of deaths to that of cases, was also retrieved.

### Modeling methods

Let  $c_{t,\Delta t}$  be weekly “reported” incidence in week  $t$  which took place in  $\Delta t$  days since the first date of corresponding week: we account for  $\Delta t$  because  $\Delta t$  greatly varied by Epidemiology Bulletin. Let  $j_t$  be the actual (unbiased) incidence in week  $t$ . That is, here we explicitly distinguish reported incidence  $c$  from actual incidence  $j$ . Setting the latest value of CFR in week 26 as the baseline (i.e. 1.0), the relative CFR of week  $t$  is denoted by  $\alpha_t$ .  $F_{\Delta t}$  is the cumulative distribution function of the time from illness onset to reporting. We assume that the actual (unbiased) CFR is a constant without any interventions or treatment and that all deaths are reported. Moreover, we assume that  $\alpha_t$  in and after week 26 remains to be the value of 1. The expected value of the weekly incidence is described as

$$E[c_{t,\Delta t}] = \frac{j_t}{\alpha_t} F_{7(s-t)+\Delta t}, \quad (1)$$

where  $s$  represents the latest week of observation (for  $t \leq s$ ).  $F_{\Delta t}$  was assumed to follow an exponential distribution with mean  $\delta$  days, i.e.,  $F_{\Delta t} = 1 - \exp(-\Delta t / \delta)$ , and thus, the product of  $j$  and  $F$  captures the reporting delay structure in empirical data.  $F_{\Delta t}$  is important in interpreting the updated epidemic curve every week. As we assume that the WHO's weekly CFR estimate mirrors the frequency of case ascertainment in each week,

$\alpha_t$  adjusts time-dependent variations in case ascertainment. The unbiased incidence,  $j_t$  in week  $t$  is obtained from the cumulative incidence,  $J(t)$ :

$$j_t = J(t) - J(t-1), \quad (2)$$

where  $J(t)$  was assumed to be described by two different parsimonious models. One is the logistic growth curve, i.e.,

$$J(t) = \frac{K}{1 + \exp(-r(t-t_i))}, \quad (3)$$

where  $K$  is referred to as the carrying capacity, representing the cumulative incidence at time infinity, i.e., our interest in the present study,  $r$  is the growth rate and  $t_i$  is the point of inflection, corresponding to the peak timing of an epidemic. The other model is a slightly more general model, referred to as the generalized logistic model or the so-called ‘‘Richards model’’, i.e.,

$$J(t) = \frac{K}{[1 + g \exp(-r(t-t_i))]^{\frac{1}{\beta}}}, \quad (4)$$

where  $g$  is the parameter that partially determines the point of inflection on vertical axis. The model (4) is known as flexible with only 4 parameters and has been widely applied to capture the temporal distribution of epidemics of variety of diseases [14–17].

Assuming that the observed data in week  $t$ ,  $c_{t,\Delta t}$  follows a Poisson distribution, the likelihood function to estimate parameters of abovementioned models (i.e., 1 parameter for  $F$  and 3 or 4 parameters for  $J$ ) is

$$L(\theta; D) = \frac{E[c_{t,\Delta t}]^{c_{t,\Delta t}} \exp(-E[c_{t,\Delta t}])}{c_{t,\Delta t}!}, \quad (5)$$

where  $\theta$  represents the population parameter and  $D$  the observed data.

The maximum likelihood estimates were obtained by minimizing the negative logarithm of (5). Profile likelihood based 95% confidence intervals (CI) were used. To compare model (3) and (4), Akaike Information Criterion (AIC) was employed.

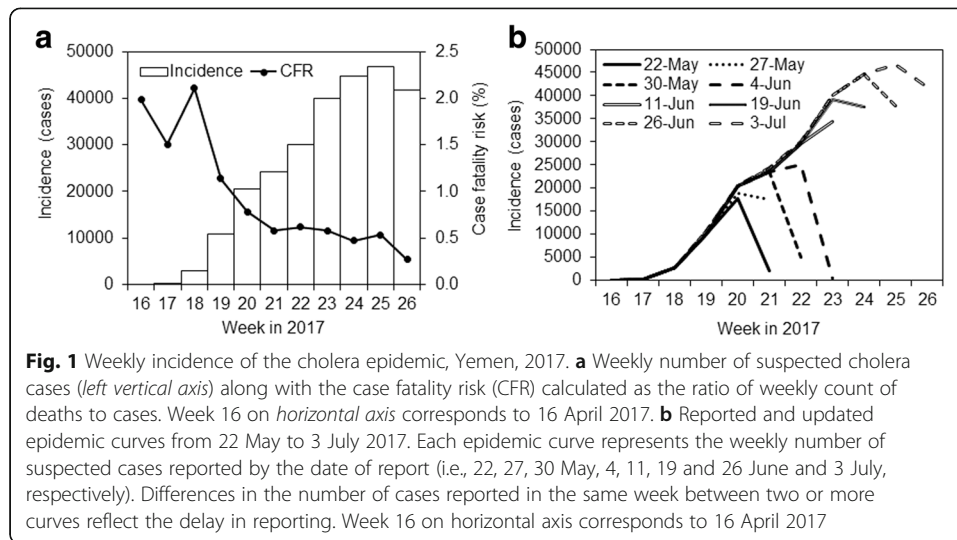
Given empirical data for each observed week, we implemented real time forecasting and sequentially updated it every week. The latest forecast that we present is based on the dataset from week 16 to 26, 2017. The forecast of unbiased cumulative number of cases was obtained from parameterized models (3) and (4). Additionally adjusting the reporting delay, we also produced forecasts of weekly incidence that is expected to be observed in week 27, 28, 29 and 30 in WHO’s report.

### Ethical considerations

The present study analyzed data that is publicly available. As such, the datasets used in our study were de-identified and fully anonymized in advance, and the analysis of publicly available data without identity information does not require ethical approval.

### Results

Figure 1 shows the epidemic curve. The incidence continuously increased from week 16, the first date of which is 16 April 2017 which we use as the initial date of the 2017 epidemic in the following analysis. Counting from 27 April 2017 (from which WHO

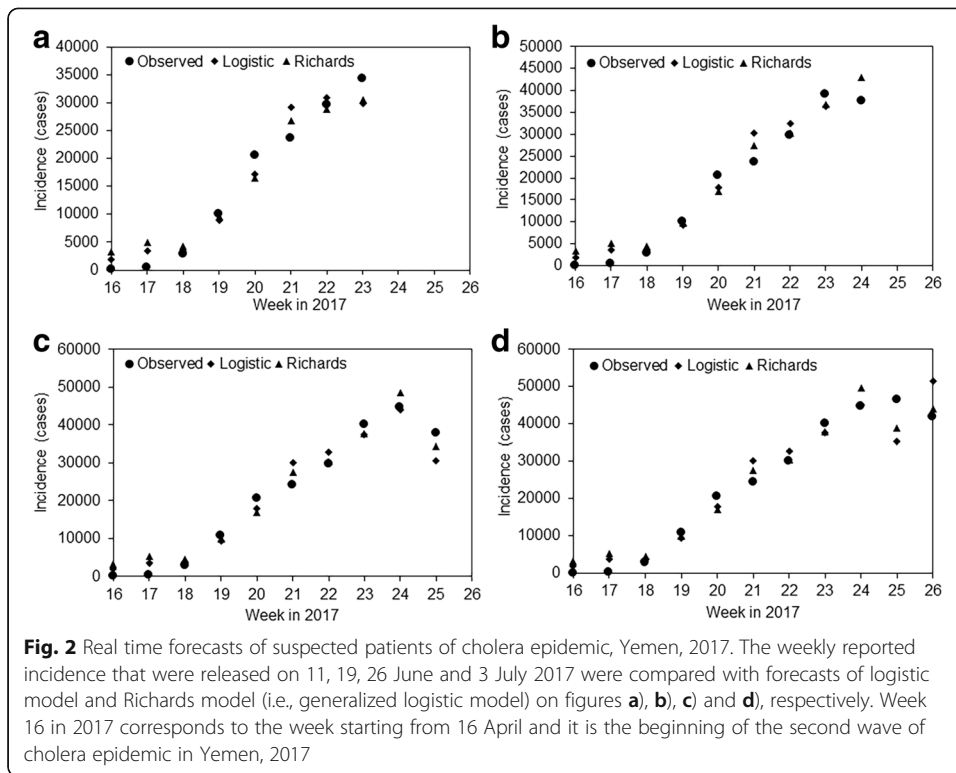


counted [12]), a total of 356,591 suspected cases and 1802 deaths were reported as of 17 July 2017. The ratio of weekly deaths to cases (i.e., the so-called CFR in the WHO's report) in the week of 25 June 2017 was 0.27, while that in the week of 16 April 2017 was 1.98. Setting the latest value of CFR (=0.27) as the baseline, those in the week of 16th, 23rd and 30th of April 2017,  $\alpha_t$ , were 7.39, 5.62 and 7.85, respectively, which we assume that they reflect small case ascertainment in April. Incidence of the latest week 26 in Fig. 1a may look indicating that the epidemic turned into decreasing trend, but Fig. 1b implicates that always the incidence in the latest week is underestimated in its initial report due to reporting delay. In the following week, the current weekly incidence likely increases.

Figure 2 shows the real-time fitting results of our model. Both logistic and generalized logistic curves qualitatively captured the observed patterns of reported incidence, and that the overall patterns of agreement have not varied with time. AIC values of using the latest data were 1,986,063 and 2,045,441 for logistic and generalized logistic models, respectively, favoring the simpler model (3). The estimated mean reporting delay were 4.4 days (95% CI: 3.2, 5.8) and 4.0 (95% CI: 2.9, 5.3) days for logistic and generalized logistic models, respectively, and the growth rate was 0.06 (95% CI: 0.05, 0.07) per day and 0.18 (95% CI: 0.09, 0.65) per day, respectively, for these models. Parameter  $g$  of Richards model was estimated at 6.9 (95% CI: 2.4, 33.2).

Figure 3 shows the forecasted  $J(t)$ , representing the actual cumulative incidence. Carrying capacity, or the cumulative incidence at the end of the epidemic, was estimated at 790,778 (95% CI: 700,495, 914,442) cases and 767,029 (95% CI: 690,877, 871,671) cases, respectively, by using logistic and generalized logistic models. Counting days from 1 January 2017 as day 0, the inflection point was estimated at day 168.0 (95% CI: 165.6, 171.1) and 173.3 (95% CI: 169.1, 180.7), respectively, for these models, indicating that the latest week (i.e. week 26) is about the time just after experiencing the inflection point of the epidemic curve.

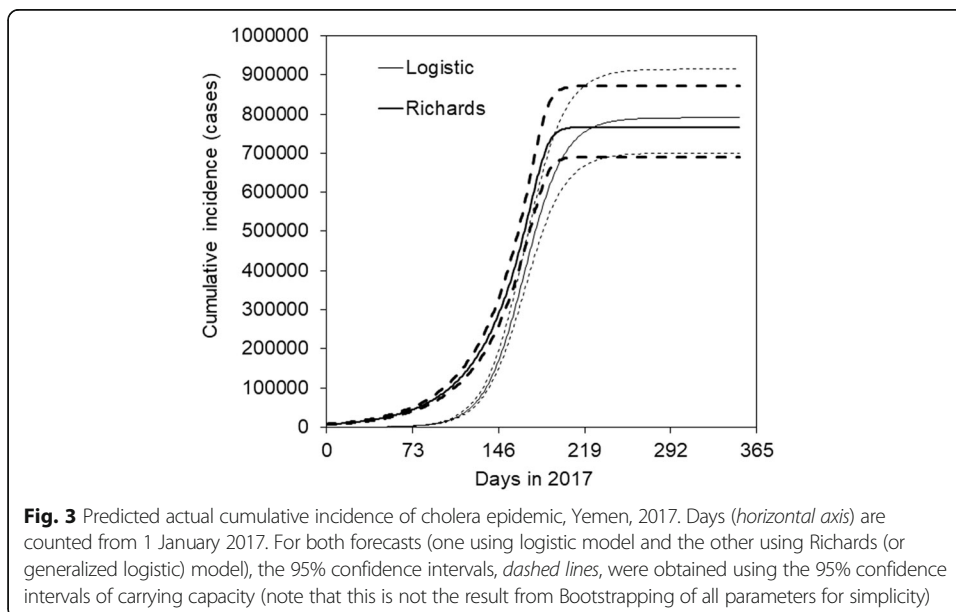
Figure 4 shows the forecasting result of observed incidence data using model (1), as derived from logistic (Fig. 4a) and generalized logistic (Fig. 4b) models. The weekly incidence that is expected to be reported in week 27 and afterwards are predicted to

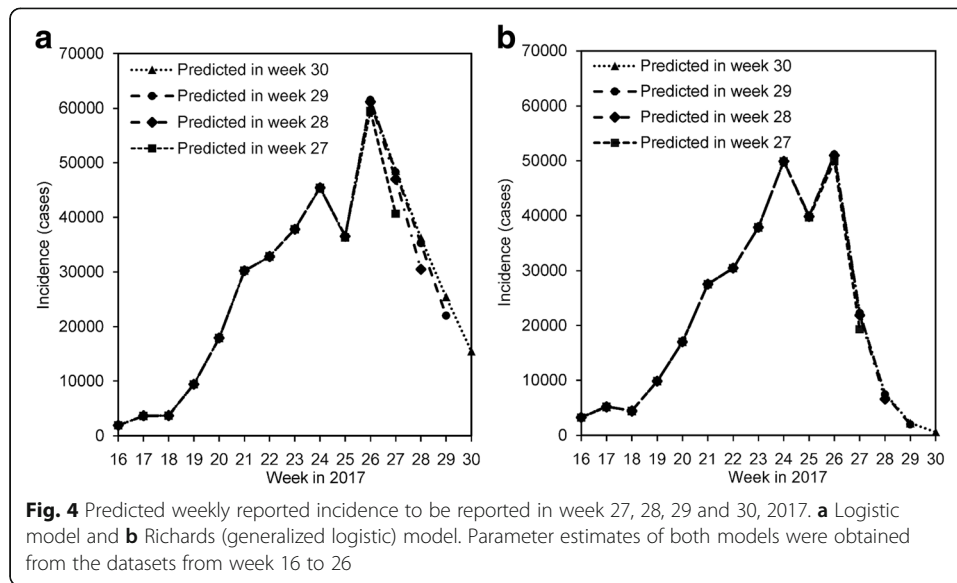


decrease over time. The logistic curve yielded slower decline with evident reporting delay compared with Richards model.

**Discussion**

The present research study responded to the cholera epidemic in Yemen, 2017 in real time. Using the weekly incidence of suspected cases, updated as a revised epidemic





curve every week, the reporting delay was explicitly incorporated into the model. Moreover, using the CFR (as calculated by the WHO which was implemented in a biased manner [18]), ascertainment bias was adjusted, enabling us to parameterize the family of logistic curves for modeling the unbiased incidence. As a result, it was estimated that we have just passed through the epidemic peak by week 26, and the unbiased cumulative incidence was forecasted to range from 690 to 910 thousand cases by the end of the epidemic. Doing so, while high incidence nearby epidemic peak may lead Yemen's people to feel that the epidemic is somewhat uncontrollable due to massive number of reported cases, we have been able to objectively demonstrate that the epidemic peak is likely over and elevate people's situation awareness through our "now-casting" approach.

To our knowledge, this modeling study is the first to elevate the situation awareness of the cholera epidemic in Yemen, 2017 via statistical forecasting of the epidemic. Employing parsimonious models with small number of parameters, the observed patterns of incidence and reporting were appropriately captured and we were successful in obtaining real time forecasts. While the epidemic peak was not directly identifiable as of week 26, 2017 only from reported data, the model has radically indicated that the inflection point has plausibly been passed. Moreover, jointly using the relative CFR with the abovementioned model with reporting delay, we assumed that the time period with high CFR involved relatively low ascertainment, enabling us to strongly feel that the qualitative patterns of unbiased epidemic curve can be captured.

Epidemiological modelers will experience similar epidemics in the future. What to be learnt from our modeling exercise are two folds. First, if the weekly incidence is reported in the up-to-the-minute manner and updated in later weeks, not a single data point but the entire epidemic curve must be precisely updated-and-reported. In the Yemen case, we initially intended to implement real time modeling using daily incidence data, but this intent was in vain mainly due to the absence of regular updates in the epidemic curves showing daily incidence. Second, not only the epidemic curve, but death data should also be consistently updated over time, so that it helps researchers to adjust ascertainment bias.

Carrying capacity, or the cumulative incidence at the end of the epidemic, is worthy of further discussions. We estimated that  $K$  ranges from 690 to 910 thousand cases, but that value is dependent on our choice of the weekly CFR in week 26 (i.e.  $\text{CFR} = 0.27$ ). If the actual CFR is greater than 0.27, our estimated  $K$  may be an overestimate. On the other hand, considering that the empirically observed CFR does not capture mild and asymptomatic infected individuals that are never reflected in the denominator of the CFR estimation, the estimated  $K$  may be a considerable underestimate.

Several limitations must be noted. First, we were not able to assure the quality of suspected cases at an individual level. Sometimes, weekly incidence in a specific week has decreased (rather than increase due to reporting delay) as the epidemic curve was updated, and in such an instance, we had to adopt the smallest latest value regardless of the week of update, because we had to avoid any decrease in the observed incidence as a function of week through minimum arbitrariness for adjustment. Second, the validity of forecast is limited due to the use of parsimonious phenomenological model. Prediction approaches using more mechanistic models would be beneficial. Third, our analysis rests on the dataset of entire Yemen, and we have not been able to dig into more detailed heterogeneous data. Spatio-temporal heterogeneity is known to play a key role in cholera transmission [19, 20] and that could allow examining the impact of environmental predictors on the transmission dynamics [21]. Fourth, our model cannot incorporate the population impact of interventions on the transmission dynamics, due to phenomenological nature of the model that we used.

While several modeling research subjects remain, the present study acted as the first to contribute to improving situation awareness of cholera epidemic in Yemen, 2017. Although the epidemic is predicted to start to decline, it is vital that ongoing resource allocations and countermeasures are tightly conducted to minimize potential number of victims.

## Conclusions

The present study responded to the cholera epidemic in Yemen, 2017 in real time. Employing parsimonious models with small number of parameters, the observed patterns of incidence and reporting were appropriately captured. It was estimated that we have just passed through the epidemic peak by the end of week 26, and the weekly incidence was predicted to start to decrease in the following weeks. The unbiased cumulative incidence was forecasted to range from 690 to 910 thousand cases by the end of the epidemic.

## Abbreviations

AIC: Akaike Information Criterion; CFR: Case fatality risk; CI: Confidence interval; WHO: World Health Organization

## Acknowledgements

Not applicable.

## Funding

HN received funding from the Japan Agency for Medical Research and Development, the Japanese Society for the Promotion of Science (KAKENHI Grant Numbers 16KT0130, 16 K15356 and 17H04701), the Japan Science and Technology Agency CREST program (JPMJCR1413) and the RISTEX program for Science of Science, Technology and Innovation Policy. BY received financial support from China Scholarship Council. The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

## Availability of data and materials

Collected datasets are publicly available from reference [12].



**Authors' contributions**

HN conceived the study. HN, ST, BY, TY and YA conceptualized the study design, collected the data, formulated mathematical model and performed statistical analyses. HN drafted the early version of the manuscript. ST, BY, TY and YA drafted figures. All authors gave comments on the revised manuscript and approved the final version of the manuscript.

**Authors' information**

The authors are experts with interest in Infectious Disease Epidemiology and also in Theoretical Epidemiology, and the team of lead author is led by professor from Hokkaido University Graduate School of Medicine.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that co-author H. Nishiura is the Editor-in-Chief of *Theoretical Biology and Medical Modelling*. This does not alter the authors' adherence to all the *Theoretical Biology and Medical Modelling* policies on sharing data and materials.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 7 July 2017 Accepted: 20 July 2017

Published online: 26 July 2017

**References**

- Kupferschmidt K. Cholera vaccine faces major test in Yemen. *Science*. 2017;356(6345):1316–7.
- Chao DL, Longini IM Jr, Morris JG Jr. Modeling cholera outbreaks. *Curr Top Microbiol Immunol*. 2014;379:195–209.
- Tuite AR, Tien J, Eisenberg M, Earn DJ, Ma J, Fisman DN. Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Ann Intern Med*. 2011;154(9):593–601.
- Andrews JR, Basu S. Transmission dynamics and control of cholera in Haiti: an epidemic model. *Lancet*. 2011;377(9773):1248–55.
- Kirpich A, Weppelmann TA, Yang Y, Ali A, Morris JG Jr, Longini IM. Cholera transmission in Ouest Department of Haiti: dynamic modeling and the future of the epidemic. *PLoS Negl Trop Dis*. 2015;9(10):e0004153.
- Mukandavire Z, Morris JG Jr. Modeling the epidemiology of cholera to prevent disease transmission in developing countries. *Microbiol Spectr*. 2015;3:3. doi:10.1128/microbiolspec.VE-0011-2014.
- Rinaldo A, Blokesch M, Bertuzzo E, Mari L, Righetto L, Murray M, Gatto M, Casagrandi R, Rodriguez-Iturbe I. A transmission model of the 2010 cholera epidemic in Haiti. *Ann Intern Med*. 2011;155(6):403–4.
- Fung IC. Cholera transmission dynamic models for public health practitioners. *Emerg Themes Epidemiol*. 2014;11(1):1.
- Fung IC, Fitter DL, Borse RH, Meltzer MI, Tappero JW. Modeling the effect of water, sanitation, and hygiene and oral cholera vaccine implementation in Haiti. *Am J Trop Med Hyg*. 2013;89(4):633–40.
- Mari L, Bertuzzo E, Finger F, Casagrandi R, Gatto M, Rinaldo A. On the predictive ability of mechanistic models for the Haitian cholera epidemic. *J R Soc Interface*. 2015;12(104):20140840.
- Abrams JY, Copeland JR, Tauxe RV, Date KA, Belay ED, Mody RK, Mintz ED. Real-time modelling used for outbreak management during a cholera epidemic, Haiti, 2010–2011. *Epidemiol Infect*. 2013;141(6):1276–85.
- World Health Organization. Cholera outbreak response. Yemen: World Health Organization; 2017. <http://www.emro.who.int/yem/yemeninfocus/situation-reports.html>. Accessed 7 July 2017
- World Health Organization. Prevention and control of cholera outbreaks: WHO policy and recommendations. Yemen: World Health Organization; 2017. <http://www.who.int/cholera/technical/prevention/control/en/index1.html>. Accessed 7 July 2017
- Dinh L, Chowell G, Mizumoto K, Nishiura H. Estimating the subcritical transmissibility of the Zika outbreak in the state of Florida, USA, 2016. *Theor Biol Med Model*. 2016;13(1):20.
- Omori R, Nishiura H. Theoretical basis to measure the impact of short-lasting control of an infectious disease on the epidemic peak. *Theor Biol Med Model*. 2011;8:2.
- Hsieh YH. Ascertaining the impact of catastrophic events on dengue outbreak: the 2014 gas explosions in Kaohsiung, Taiwan. *PLoS One*. 2017;12(5):e0177422.
- Hsieh YH, Ma S. Intervention measures, turning point, and reproduction number for dengue, Singapore, 2005. *Am J Trop Med Hyg*. 2009;80(1):66–71.
- Ejima K, Omori R, Cowling BJ, Aihara K, Nishiura H. The time required to estimate the case fatality ratio of influenza using only the tip of an iceberg: joint estimation of the virulence and the transmission potential. *Comput Math Methods Med*. 2012;2012:978901.
- Azman AS, Luquero FJ, Rodrigues A, Palma PP, Grais RF, Banga CN, Grenfell BT, Lessler J. Urban cholera transmission hotspots and their implications for reactive vaccination: evidence from Bissau city, Guinea Bissau. *PLoS Negl Trop Dis*. 2012;6(11):e1901.
- Kelly MR Jr, Tien JH, Eisenberg MC, Lenhart S. The impact of spatial arrangements on epidemic disease dynamics and intervention strategies. *J Biol Dyn*. 2016;10:222–49.
- Eisenberg MC, Kujbida G, Tuite AR, Fisman DN, Tien JH. Examining rainfall and cholera dynamics in Haiti using statistical and dynamic modeling approaches. *Epidemics*. 2013;5:197–207.