2012

A Generalized Cholera Model and Epidemic-Endemic Analysis

Jin Wang
Old Dominion University

Shu Liao

Original Publication Citation

Wang, J., & Liao, S. (2012). A generalized cholera model and epidemic-endemic analysis. *Journal of Biological Dynamics, 6*(2), 568-589. doi:10.1080/17513758.2012.658089

Journal of Biological Dynamics
Vol. 6, No. 2, March 2012, 568–589

Taylor & Francis
Taylor & Francis Group

A generalized cholera model and epidemic–endemic analysis

Jin Wang[a]* and Shu Liao[b]

[a]*Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA;* [b]*School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China*

(*Received 9 May 2011; final version received 6 January 2012*)

The transmission of cholera involves both human-to-human and environment-to-human pathways that complicate its dynamics. In this paper, we present a new and unified deterministic model that incorporates a general incidence rate and a general formulation of the pathogen concentration to analyse the dynamics of cholera. Particularly, this work unifies many existing cholera models proposed by different authors. We conduct equilibrium analysis to carefully study the complex epidemic and endemic behaviour of the disease. Our results show that despite the incorporation of the environmental component, there exists a forward transcritical bifurcation at $R_0 = 1$ for the combined human–environment epidemiological model under biologically reasonable conditions.

Keywords: cholera model; stability; epidemic and endemic dynamics

1. Introduction

Despite many clinical and theoretical studies [1,14,19,21,38,41,42,52] and tremendous administrative efforts and interventions,[1] cholera remains a significant threat to public health in developing countries. In the year 2006 alone, about 240,000 cholera cases were officially notified to the World Health Organization (WHO), with Africa accounting for the majority of these cases. Recent cholera outbreaks in Haiti (2010–2011), Nigeria (2010), Kenya (2010), Vietnam (2009), Zimbabwe (2008–2009), Iraq (2008), Congo (2008), and India (2007) continue leading to a large number of infections and deaths and receiving worldwide attention.[1,2,3]

Cholera is an acute intestinal infection caused by the bacterium *Vibrio cholerae*. Its dynamics are complicated by the multiple interactions between the human host, the pathogen and the environment [41], which contribute to both direct human-to-human and indirect environment-to-human transmission pathways. A deep understanding of the disease dynamics would provide important guidelines to the effective prevention and control strategies [6,16]. Mathematical modelling, simulation and analysis offer a promising way to look into the nature of cholera dynamics, and many efforts have been devoted to this topic. Below, we briefly review some representative mathematical models proposed by various authors.

*Corresponding author. E-mail: j3wang@odu.edu
Author email: ywm2006@yahoo.com

ISSN 1751-3758 print/ISSN 1751-3766 online

http://dx.doi.org/10.1080/17513758.2012.658089
http://www.tandfonline.com

Capasso and Paveri-Fontana [2] introduced a simple deterministic model in 1979 to study a cholera epidemic in the Mediterranean. They considered a population of bacteria and a population of infected humans, with infectivity modelled under a saturation condition. Pourabbas *et al.* published an SIRS cholera model [43] in 2001, representing human-to-human transmission with a time-dependent infectivity coefficient $\beta(t)$, so that the incidence rate was given by $\beta(t)I$ (where I denotes the infected). Also in 2001, Codeço proposed a model [5] that explicitly accounted for the environmental component, i.e., the *V. cholerae* concentration in the water supply, denoted as B, into a regular SIR epidemiological model. The incidence was modelled by a logistic function, $a(B/(K+B))$, where a is the contact rate with contaminated water and K is the half-saturation rate (i.e., ID_{50}, the infectious dose in water sufficient to produce disease in 50% of those exposed). Ghosh *et al.* in 2004 published an SIS model [11], which included both the concentration B of vibrios and the density E of environmental discharge that contributes to the growth of the vibrio population. The model had both human-to-human and environment-to-human transmission modes with the incidence given by $\beta I + \lambda B$, where β and λ are corresponding contact rates. In 2006, Hartley *et al*. [12] extended Codeço's work to include a hyperinfectious state of the pathogen, representing the 'explosive' infectivity of freshly shed *V. cholerae*, based on the laboratory measurements that freshly shed *V. cholerae* from human intestines outcompeted other *V. cholerae* by as much as 700-fold for the first few hours in the environment [1,38]. They modelled the incidence factor by $\beta_{\mathrm{L}}(B_{\mathrm{L}}/(\kappa_{\mathrm{L}}+B_{\mathrm{L}})) + \beta_{\mathrm{H}}(B_{\mathrm{H}}/(\kappa_{\mathrm{H}}+B_{\mathrm{H}}))$, where β_{H} and β_{L} are the hyperinfectious (HI) and less-infectious (LI) ingestion rates, and κ_{H} and κ_{L} are the HI and LI half-saturation rates. This model was also analysed by Liao and Wang [33]. Most recently, Mukandavire *et al.* [40] proposed a model to study the 2008–2009 cholera outbreak in Zimbabwe. The model considered both human-to-human and environment-to-human transmission pathways; the incidence factor was represented by $\beta_{\mathrm{e}}(B/(K+B)) + \beta_{\mathrm{h}}I$ with β_{e} and β_{h} being the rates of vibrio ingestion from the environment and the human–human interaction, respectively. This work demonstrated the importance of the human-to-human transmission in cholera epidemics, especially in places such as Zimbabwe, a land-locked country in the middle of Africa. In addition, Tien and Earn [51] in 2010 published a water-borne disease model which also included the dual-transmission pathways, with bilinear incidence rates employed for both the environment-to-human and human-to-human infection routes. No saturation effect was considered in the work of Tien and Earn.

All these models have their own strength and weakness; some models only track the human population dynamics directly and represent the *V. cholerae* population in a separate manner, whereas some other models tend to focus on the environmental components while neglecting direct human-to-human transmission. The dynamics of bacterial growth in water are assumed to be linear in all the aforementioned works. Meanwhile, many of these studies lack rigorous mathematical analysis which accounts for part of the reason that cholera dynamics has not been well understood so far. Indeed, the interaction between *V. cholerae* and susceptible human population could be more complicated than being linear or logistic. Furthermore, the bacterial growth outside of human hosts does not have to follow linear dynamics. For example, Jensen *et al.* [18] proposed a mathematical model to investigate the control of cholera outbreaks with respect to bacteriophage. The incidence was modelled as $\pi(B/(Ck+B))^7$, a highly nonlinear function. The growth of *V. cholerae* is also nonlinear (a quadratic function in B) in their model.

Two major differences among these models, as mentioned above, are how the incidence rate is determined and how the environmental vibrio concentration is formulated. Hence, the goal of the present paper is to propose a unified cholera model that allows general nonlinear incidence factors and general representation of the pathogen concentrations. Based on this general model, we will conduct a careful mathematical study to explore the complex cholera dynamics so as to improve our understanding of the fundamental disease transmission mechanism. We will particularly investigate the stability property in both the epidemic and endemic dynamics through

equilibrium analysis. Within this general framework, we will be able to unify existing models in the analysis and simulation.

The remainder of this paper is organized as follows. In Section 2, we introduce the generalized model and state the necessary assumptions. In Sections 3, we derive the basic reproduction number using the next-generation matrix approach, followed by the global stability analysis of the disease-free equilibrium (DFE) in Section 4. We show the existence and uniqueness of the endemic equilibrium in Section 5, and conduct the endemic stability analysis in Section 6. We then briefly study in Section 7 several existing models as special cases of our proposed framework. Finally, we draw conclusion and provide discussion in Section 8.

2. Model and notations

We construct the following differential equations for the cholera dynamics based on the combination of a regular SIR model and an environmental component:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = bN - Sf(I,B) - bS, \tag{1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = Sf(I,B) - (\gamma + b)I, \tag{2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - bR, \tag{3}$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = h(I,B), \tag{4}$$

where, as usual, S, I, and R denote the susceptible, the infected, and the recovered populations, respectively, and B denotes the concentration of the vibrios in the contaminated water. The total population $N = S + I + R$ is assumed to be a constant. The parameter b represents the natural human birth/death rate, and γ represents the rate of recovery from cholera. In this generalized model, $f(I,B)$ is the incidence function that determines the rate of new infection. For example, $f(I,B) = a(B/(K+B))$ in Codeço's model [5], $f(I,B) = \beta I + \lambda B$ in the model of Ghosh *et al.* [11] and $f(I,B) = \beta_{\mathrm{e}}(B/(K+B)) + \beta_{\mathrm{h}} I$ in the model of Mukandavire *et al.* [40]. Finally, the function $h(I,B)$ describes the rate of change for the pathogen in the environment which can be either linear or nonlinear.

If we set

$$X = (S, I, R, B)^{\mathrm{T}}, \tag{5}$$

then the above equations can be put in a vector form as

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \mathbf{F}(X). \tag{6}$$

Remark 2.1 We allow B to be either a scalar or a vector in this system in order to facilitate more general formulation. For example, if we consider both the HI and LI states of the vibrios, then we may write $B = [B_{\mathrm{H}}, B_{\mathrm{L}}]$. In such a case, it is understood that

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \begin{bmatrix} \dfrac{\mathrm{d}B_{\mathrm{H}}}{\mathrm{d}t} \\ \dfrac{\mathrm{d}B_{\mathrm{L}}}{\mathrm{d}t} \end{bmatrix}, \quad h(I,B) = \begin{bmatrix} h_{\mathrm{H}}(I,B) \\ h_{\mathrm{L}}(I,B) \end{bmatrix}$$

and

$$\frac{\partial f}{\partial B} = \begin{bmatrix} \frac{\partial f}{\partial B_{\mathrm{H}}} \\ \frac{\partial f}{\partial B_{\mathrm{L}}} \end{bmatrix}, \quad \frac{\partial^2 f}{\partial B^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial B_{\mathrm{H}}^2} & \frac{\partial^2 f}{\partial B_{\mathrm{H}} \partial B_{\mathrm{L}}} \\ \frac{\partial^2 f}{\partial B_{\mathrm{L}} \partial B_{\mathrm{H}}} & \frac{\partial^2 f}{\partial B_{\mathrm{L}}^2} \end{bmatrix}, \quad \frac{\partial h}{\partial B} = \begin{bmatrix} \frac{\partial h_{\mathrm{H}}}{\partial B_{\mathrm{H}}} & \frac{\partial h_{\mathrm{H}}}{\partial B_{\mathrm{L}}} \\ \frac{\partial h_{\mathrm{L}}}{\partial B_{\mathrm{H}}} & \frac{\partial h_{\mathrm{L}}}{\partial B_{\mathrm{L}}} \end{bmatrix}, \quad \text{etc.}$$

Remark 2.2 We write a vector $V \geq 0$ (≤ 0) if each component of V is ≥ 0 (≤ 0). We write a matrix $A \geq 0$ (≤ 0) if A is positive (negative) semidefinite.

To make biological sense for our model, we assume that the two functions f and h satisfy the following conditions for $I \geq 0, B \geq 0$:

(a)

$$f(0,0) = 0, \quad h(0,0) = 0,$$

(b)

$$f(I,B) \geq 0,$$

(c)

$$\frac{\partial f}{\partial I}(I,B) \geq 0, \quad \frac{\partial f}{\partial B}(I,B) \geq 0, \quad \frac{\partial h}{\partial I}(I,B) \geq 0, \quad \frac{\partial h}{\partial B}(I,B) \leq 0,$$

(d) $f(I,B)$ is concave; i.e., the matrix

$$D^2 f \triangleq \begin{bmatrix} \frac{\partial^2 f}{\partial I^2} & \frac{\partial^2 f}{\partial I \partial B} \\ \frac{\partial^2 f}{\partial B \partial I} & \frac{\partial^2 f}{\partial B^2} \end{bmatrix}$$

is negative semidefinite everywhere.

(e) $h(I,B)$ is concave; i.e., the matrix $D^2 h$ is negative semidefinite everywhere.

The assumption (a) ensures the existence of a unique DFE for system (6), i.e.,

$$X_0 = (N,0,0,0)^{\mathrm{T}}. \tag{7}$$

The assumption (b) ensures a positive incidence rate. The first two inequalities in assumption (c) state that increased infection and pathogen concentration lead to higher incidence rate (owing to higher level of transmission), whereas the third inequality states that increased infection results in higher growth rate for the pathogen in the environment (owing to higher shedding rate). The last inequality in assumption (c) indicates a positive net death rate of the vibrios. Condition (d) is a common assumption for nonlinear incidence [17,24,39]. In our model, this condition regulates $f(I,B)$ as a biologically realistic incidence based on a consequence of saturation effects: when the number of the infective, or the environmental pathogen concentration, is high, the incidence rate will respond more slowly than linearly to the increase in I and B. Similar arguments lead to the condition (e), which is an additional assumption we introduce for the regulation of the environmental function $h(I,B)$.

Furthermore, we assume that the equation $h(I,B) = 0$ implicitly defines a function $B = g(I)$, which satisfies the following condition:

(f)

$$g'(I) \geq 0, \quad g''(I) \leq 0, \quad \text{for } I \geq 0.$$

This assumption states that the pathogen concentration increases with the number of the infected, while the rate of increase will be below linear when the infected population is high owing to saturation effects.

Based on the assumption (b), it is straightforward to see that if any component of (S, I, R) becomes 0, then the derivative of this component will be non-negative. Meanwhile, as $(\mathrm{d}/\mathrm{d}t)(S + I + R) = 0$, $S(t) + I(t) + R(t)$ will remain a constant (i.e., N) for all $t \geq 0$. Hence, the following result can be easily established.

LEMMA 2.3 *If* $S(0) \geq 0, I(0) \geq 0, R(0) \geq 0$ *and* $S(0) + I(0) + R(0) = N$, *then* $S(t) \geq 0, I(t) \geq 0, R(t) \geq 0$ *and* $S(t) + I(t) + R(t) = N$, *for all* $t \geq 0$.

Remark 2.4 Lemma 2.3 ensures that the solution of the model system (1)–(4) is biologically feasible for all times. Mathematically speaking, the solution domain

$$\bar{D} = \{(S, I, R) \mid S \geq 0,\ I \geq 0,\ R \geq 0,\ S + I + R = N\}$$

is a positively invariant set in $\mathbb{R}^3$.

3. Next-generation matrix analysis

We start our analysis by determining the basic reproduction number, R_0, of our proposed model. Based on the work of van den Driessche and Watmough [53], R_0 is mathematically defined as the spectral radius of the next-generation matrix. In our system (1)–(4), only I and B are directly related to the infection. Following the approach of van den Driessche and Watmough [53], we write

$$\begin{bmatrix} \dfrac{\mathrm{d}I}{\mathrm{d}t} \\ \dfrac{\mathrm{d}B}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} Sf(I,B) \\ 0 \end{bmatrix} - \begin{bmatrix} (\gamma + b)I \\ -h(I,B) \end{bmatrix} = \mathcal{F} - \mathcal{V}, \tag{8}$$

where $\mathcal{F}$ denotes the rate of appearance of new infections, and $\mathcal{V}$ denotes the rate of transfer of individuals into or out of each population set.

The next-generation matrix is defined as FV^{-1}, where F and V are 2×2 Jacobian matrices given by

$$F = D\mathcal{F}(X_0) = \begin{bmatrix} N\dfrac{\partial f}{\partial I}(0,0) & N\dfrac{\partial f}{\partial B}(0,0) \\ 0 & 0 \end{bmatrix}, \quad V = D\mathcal{V}(X_0) = \begin{bmatrix} \gamma + b & 0 \\ -\dfrac{\partial h}{\partial I}(0,0) & -\dfrac{\partial h}{\partial B}(0,0) \end{bmatrix},$$

where X_0 is the DFE defined in Equation (7). After some algebra, we obtain

$$V^{-1} = \frac{-1}{\gamma + b} \begin{bmatrix} -1 & 0 \\ \left(\dfrac{\partial h}{\partial B}(0,0)\right)^{-1} \dfrac{\partial h}{\partial I}(0,0) & \left(\dfrac{\partial h}{\partial B}(0,0)\right)^{-1} (\gamma + b) \end{bmatrix}$$

Hence, the next-generation matrix is

$$FV^{-1} = \frac{-1}{\gamma + b} \begin{bmatrix} -N \begin{bmatrix} \dfrac{\partial f}{\partial I}(0,0) - \dfrac{\partial f}{\partial B}(0,0) \\ \left(\dfrac{\partial h}{\partial B}(0,0)\right)^{-1} \dfrac{\partial h}{\partial I}(0,0) \end{bmatrix} & N(\gamma + b) \left(\dfrac{\partial h}{\partial B}(0,0)\right)^{-1} \dfrac{\partial f}{\partial B}(0,0) \\ 0 & 0 \end{bmatrix}.$$

Its spectral radius $\rho(FV^{-1}) = \max_{1\le i\le 2}|\lambda_i|$, where λ_i denotes the ith eigenvalue, can be easily found. Therefore, we obtain the basic reproduction number as

$$R_0 = \frac{N}{\gamma + b}\left[\frac{\partial f}{\partial I}(0,0) - \frac{\partial f}{\partial B}(0,0)\left(\frac{\partial h}{\partial B}(0,0)\right)^{-1}\frac{\partial h}{\partial I}(0,0)\right]. \tag{9}$$

By our assumption, $h(I,B) = 0$ defines an implicit function $B = g(I)$ with $g'(I) \ge 0$. Using implicit differentiation, we obtain $\partial h/\partial I + (\partial h/\partial B)g'(I) = 0$, which yields

$$g'(I) = -\left(\frac{\partial h}{\partial B}\right)^{-1}\frac{\partial h}{\partial I} \tag{10}$$

for $I \ge 0$. Substituting Equation (10) into Equation (9), we obtain

$$R_0 = \frac{N}{\gamma + b}\frac{\partial f}{\partial I}(0,0) + \frac{N}{\gamma + b}\frac{\partial f}{\partial B}(0,0)g'(0) \triangleq R_0^{\mathrm{hh}} + R_0^{\mathrm{eh}}. \tag{11}$$

Equation (11) clearly shows that R_0 depends on two factors: one is due to human-to-human transmission (R_0^{hh}) and the other is due to environment-to-human transmission (R_0^{eh}). If $(\partial f/\partial I)(0,0) = 0$, then $R_0 = R_0^{\mathrm{eh}}$; if $(\partial f/\partial B)(0,0) = 0$, then $R_0 = R_0^{\mathrm{hh}}$. In general, both R_0^{hh} and R_0^{eh} contribute to the basic reproduction rate. Biologically speaking, R_0 measures the average number of secondary infections that occur when one infective is introduced into a completely susceptible host population [16,53,54]. In Equation (11), the term $1/(\gamma + b)$ represents the expected time of the infection, $(\partial f/\partial I)(0,0)$ represents the unit human-to-human transmission rate and $(N/(\gamma + b))(\partial f/\partial I)(0,0)$ measures the total number of secondary infections caused by the human-to-human transmission. Similarly, the product $(\partial f/\partial B)(0,0)g'(0)$ represents the unit environment-to-human transmission rate, and $(N/(\gamma + b))(\partial f/\partial B)(0,0)g'(0)$ measures the total number of secondary infections caused by the environment-to-human transmission.

Remark 3.1 It can be easily verified that this derivation of R_0 holds true no matter B is a scalar or vector. In case B is a vector, $(\partial h/\partial B)(0,0)$ is a matrix (see Remark 2.1) and $((\partial h/\partial B)(0,0))^{-1}$ represents its inverse.

Based on the framework explained by van den Driessche and Watmough [53], we immediately obtain the result below regarding the local asymptotical stability of the DFE.

THEOREM 3.2 *Let R_0 be defined in Equation* (11). *The DFE of the system* (1)–(4) *is locally asymptotically stable if $R_0 < 1$, and unstable if $R_0 > 1$.*

4. Global stability of DFE

To study the global asymptotic stability of the DFE, we need the following result introduced by Castillo-Chavez *et al.* [3].

LEMMA 4.1 [3] *Consider a model system written in the form*

$$\begin{aligned}\frac{\mathrm{d}X_1}{\mathrm{d}t} &= F(X_1, X_2),\\ \frac{\mathrm{d}X_2}{\mathrm{d}t} &= G(X_1, X_2), \quad G(X_1, 0) = 0,\end{aligned} \tag{12}$$

where $X_1 \in \mathbb{R}^m$ denotes (*its components*) *the number of uninfected individuals and $X_2 \in \mathbb{R}^n$ denotes* (*its components*) *the number of infected individuals including latent, infectious, etc*; *$X_0 = (X_1^*, 0)$ denotes the DFE of the system.*

Also assume the conditions (H1) *and* (H2) *below*:

(H1) *For* $\mathrm{d}X_1/\mathrm{d}t = F(X_1, 0)$, X_1^* *is globally asymptotically stable*;
(H2) $G(X_1, X_2) = AX_2 - \hat{G}(X_1, X_2)$, $\hat{G}(X_1, X_2) \geq 0$ *for* $(X_1, X_2) \in \Omega$, *where the Jacobian* $A = (\partial G/\partial X_2)(X_1^*, 0)$ *is an M-matrix* (*the off-diagonal elements of A are non-negative*) *and* Ω *is the region where the model makes biological sense.*

Then the DFE $X_0 = (X_1^*, 0)$ *is globally asymptotically stable provided that* $R_0 < 1$.

THEOREM 4.2 *The DFE of the model* (1)–(4) *is globally asymptotically stable if* $R_0 < 1$.

Proof We adopt the notations in Lemma 4.1 and verify the conditions (H1) and (H2). In our model, $X_1 = (S, R)^{\mathrm{T}}$, $X_2 = (I, B)^{\mathrm{T}}$ and $X_1^* = (N, 0)^{\mathrm{T}}$. The uninfected subsystem is

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} S \\ R \end{bmatrix} = F = \begin{bmatrix} bN - bS - Sf(I, B) \\ \gamma I - bR \end{bmatrix} \tag{13}$$

and the infected subsystem is

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} I \\ B \end{bmatrix} = G = \begin{bmatrix} Sf(I, B) - (\gamma + b)I \\ h(I, B) \end{bmatrix} \tag{14}$$

When $I = B = 0$ (i.e., $X_2 = 0$), the uninfected subsystem (13) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} S \\ R \end{bmatrix} = \begin{bmatrix} bN - bS \\ -bR \end{bmatrix} \tag{15}$$

and its solution is

$$R(t) = R(0)\,\mathrm{e}^{-bt}, \quad S(t) = N - (N - S(0))\,\mathrm{e}^{-bt}.$$

Clearly, $R(t) \to 0$ and $S(t) \to N$ as $t \to \infty$, regardless of the values of $R(0)$ and $S(0)$. Hence, $X_1^* = (N, 0)$ is globally asymptotically stable for the subsystem

$$\frac{\mathrm{d}X_1}{\mathrm{d}t} = F(X_1, 0).$$

Next, we have

$$\begin{aligned} G &= \frac{\partial G}{\partial X_2}(N, 0, 0, 0)X_2 - \hat{G} \\ &= \begin{bmatrix} N\dfrac{\partial f}{\partial I}(0, 0) - (\gamma + b) & N\dfrac{\partial f}{\partial B}(0, 0) \\ \dfrac{\partial h}{\partial I}(0, 0) & \dfrac{\partial h}{\partial B}(0, 0) \end{bmatrix} \begin{bmatrix} I \\ B \end{bmatrix} - \begin{bmatrix} N\dfrac{\partial f}{\partial I}(0, 0)I + N\dfrac{\partial f}{\partial B}(0, 0)B - Sf(I, B) \\ \dfrac{\partial h}{\partial I}(0, 0)I + \dfrac{\partial h}{\partial B}(0, 0)B - h(I, B) \end{bmatrix}. \end{aligned}$$

Obviously, $A = (\partial G/\partial X_2)(N, 0, 0, 0)$ is an M-matrix based on the assumption (c). It remains to show $\hat{G} \geq 0$. The assumption (d) implies that the surface $f = f(I, B)$ is below its tangent plane at

any point $(I_0, B_0) \geq 0$; that is,

$$f(I,B) \leq f(I_0,B_0) + \frac{\partial f}{\partial I}(I_0,B_0)(I-I_0) + \frac{\partial f}{\partial B}(I_0,B_0)(B-B_0). \tag{16}$$

Particularly, setting $(I_0, B_0) = (0, 0)$ and using the assumption (a), we obtain

$$f(I,B) \leq \frac{\partial f}{\partial I}(0,0)I + \frac{\partial f}{\partial B}(0,0)B, \tag{17}$$

for all $(I, B) \geq 0$. A similar argument leads to, according to the assumption (e),

$$h(I,B) \leq \frac{\partial h}{\partial I}(0,0)I + \frac{\partial h}{\partial B}(0,0)B. \tag{18}$$

Hence, $\hat{G} \geq 0$ for $I \geq 0, B \geq 0$.

Based on Lemma 4.1, the DFE $X_0 = (N, 0, 0, 0)$ is globally asymptotically stable when $R_0 < 1$. ■

COROLLARY 4.3 *If* $R_0 < 1$, *then* $\lim_{t\to\infty} X(t) = X_0$ *for any solution* $X(t)$ *of the system* (1)–(4).

5. The endemic equilibrium

Theorem 4.2 completely determines the global dynamics of our model when $R_0 < 1$. The epidemiological consequence is that the number of the infected, no matter how large initially, will vanish in time so that the disease dies out. In contrast, the disease will persist when $R_0 > 1$. To investigate the resulted long-term dynamics, we turn to the endemic analysis in what follows.

The theorem below shows the existence and uniqueness of the endemic equilibrium.

THEOREM 5.1 *For the model system* (1)–(4), *there exists a unique positive endemic equilibrium if* $R_0 > 1$, *and there is no positive endemic equilibrium if* $R_0 < 1$.

Proof By our assumption, $h(I, B) = 0$ defines an implicit function $B = g(I)$. Meanwhile, by setting the right-hand sides of Equations (1) and (2) to zero, we obtain

$$S = \frac{bN}{b + f(I, g(I))}, \quad I = \frac{Sf(I, g(I))}{(\gamma + b)}, \tag{19}$$

which yields

$$I = \tilde{H}(I) \triangleq \frac{bNf(I, g(I))}{(\gamma + b)[b + f(I, g(I))]}. \tag{20}$$

Now the question is whether $\tilde{H}(I)$ has a non-trivial fixed point on $(0, \infty)$.

Clearly, $\tilde{H}(I) \geq 0$ for $I \geq 0$, and $\tilde{H}(0) = 0$. Let us denote $P(I) = f(I, g(I))$. Then

$$\begin{aligned}\tilde{H}'(I) &= \frac{bN}{\gamma + b} \frac{(b + P(I))P'(I) - P(I)P'(I)}{[b + P(I)]^2} \\ &= \frac{bN}{\gamma + b} \frac{bP'(I)}{[b + P(I)]^2},\end{aligned} \tag{21}$$

where

$$P'(I) = \frac{\partial f}{\partial I} + \frac{\partial f}{\partial B} g'(I) \geq 0 \tag{22}$$

due to assumptions (c) and (f). Thus, $\tilde{H}'(I) \geq 0$ for $I \geq 0$. In particular,

$$\tilde{H}'(0) = \frac{N}{\gamma + b} P'(0) = R_0. \tag{23}$$

Next, we have

$$\tilde{H}''(I) = \frac{b^2 N}{(\gamma + b)[b + P(I)]^3} [(b + P(I))P''(I) - 2(P'(I))^2], \tag{24}$$

where

$$\begin{aligned}P''(I) &= \frac{\partial^2 f}{\partial I^2} + 2g'(I) \frac{\partial^2 f}{\partial I \partial B} + (g'(I))^2 \frac{\partial^2 f}{\partial B^2} + \frac{\partial f}{\partial B} g''(I) \\ &= \begin{bmatrix} 1, g'(I) \end{bmatrix} \begin{bmatrix} \dfrac{\partial^2 f}{\partial I^2} & \dfrac{\partial^2 f}{\partial I \partial B} \\ \dfrac{\partial^2 f}{\partial I \partial B} & \dfrac{\partial^2 f}{\partial B^2} \end{bmatrix} \begin{bmatrix} 1 \\ g'(I) \end{bmatrix} + \frac{\partial f}{\partial B} g''(I).\end{aligned} \tag{25}$$

Based on the assumption (d), the matrix

$$\begin{bmatrix} \dfrac{\partial^2 f}{\partial I^2} & \dfrac{\partial^2 f}{\partial I \partial B} \\ \dfrac{\partial^2 f}{\partial I \partial B} & \dfrac{\partial^2 f}{\partial B^2} \end{bmatrix}$$

is negative semidefinite. Meanwhile, $g''(I) \leq 0$ due to the assumption (f). Thus, $P''(I) \leq 0$. Consequently, $\tilde{H}''(I) \leq 0$ for all $I \geq 0$. Therefore, $\tilde{H}(I)$ is increasing and concave on $(0, \infty)$ with $\tilde{H}(0) = 0$.

Clearly, if $\tilde{H}'(0) = R_0 > 1$, there is a unique positive fixed point I^* for $\tilde{H}$ (Figure 1(a)). If $\tilde{H}'(0) = R_0 < 1$, there is no positive fixed point for $\tilde{H}$ (Figure 1(b)). ■

Remark 5.2 The same result that holds in case B is a vector, say, $B = [B_{\mathrm{H}}, B_{\mathrm{L}}]$. In fact, we can write $g(I) = [g_{\mathrm{H}}(I),\ g_{\mathrm{L}}(I)]^{\mathrm{T}}$ and $g'(I) = [g'_{\mathrm{H}}(I),\ g'_{\mathrm{L}}(I)]^{\mathrm{T}}$. It is then straightforward to verify that

$$P'(I) = \frac{\partial f}{\partial I} + \begin{bmatrix} \dfrac{\partial f}{\partial B_{\mathrm{H}}}, & \dfrac{\partial f}{\partial B_{\mathrm{L}}} \end{bmatrix} \begin{bmatrix} g'_{\mathrm{H}} \\ g'_{\mathrm{L}} \end{bmatrix} = \frac{\partial f}{\partial I} + \frac{\partial f}{\partial B} g'(I) \geq 0,$$

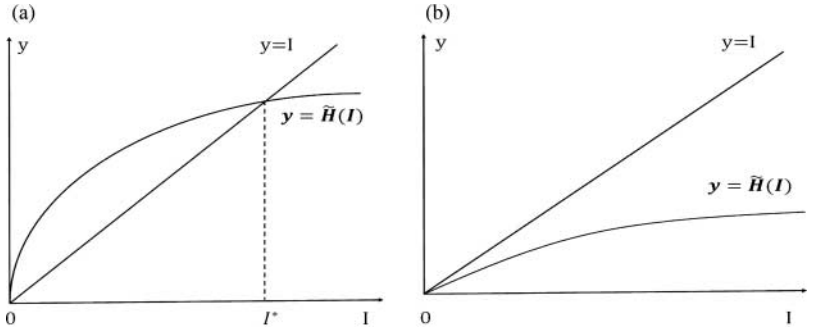

Figure 1. Two typical scenarios for the function $\tilde{H}(I)$ defined in Equation (20): (a) when $\tilde{H}'(0) > 1$, the curve $y = \tilde{H}(I)$ has a unique intersection with the line $y = I$ for $I > 0$; and (b) when $\tilde{H}'(0) < 1$, the curve $y = \tilde{H}(I)$ has no intersection with the line $y = I$ for $I > 0$.

which takes the same form as (22). Meanwhile,

$$\begin{aligned} P''(I) &= \frac{\partial^2 f}{\partial I^2} + 2\left[\frac{\partial^2 f}{\partial I \partial B_{\mathrm{H}}}, \quad \frac{\partial^2 f}{\partial I \partial B_{\mathrm{L}}}\right]\begin{bmatrix} g'_{\mathrm{H}} \\ g'_{\mathrm{L}} \end{bmatrix} \\ &\quad + [g'_{\mathrm{H}},\ g'_{\mathrm{L}}]\begin{bmatrix} \dfrac{\partial^2 f}{\partial B_{\mathrm{H}}^2} & \dfrac{\partial^2 f}{\partial B_{\mathrm{H}} \partial B_{\mathrm{L}}} \\ \dfrac{\partial^2 f}{\partial B_{\mathrm{L}} \partial B_{\mathrm{H}}} & \dfrac{\partial^2 f}{\partial B_{\mathrm{L}}^2} \end{bmatrix}\begin{bmatrix} g'_{\mathrm{H}} \\ g'_{\mathrm{L}} \end{bmatrix} + \left[\frac{\partial f}{\partial B_{\mathrm{H}}}, \ \frac{\partial f}{\partial B_{\mathrm{L}}}\right]\begin{bmatrix} g''_{\mathrm{H}} \\ g''_{\mathrm{L}} \end{bmatrix} \\ &= [1, g'_{\mathrm{H}}, g'_{\mathrm{L}}]\begin{bmatrix} \dfrac{\partial^2 f}{\partial I^2} & \dfrac{\partial^2 f}{\partial I \partial B_{\mathrm{H}}} & \dfrac{\partial^2 f}{\partial I \partial B_{\mathrm{L}}} \\ \dfrac{\partial^2 f}{\partial I \partial B_{\mathrm{H}}} & \dfrac{\partial^2 f}{\partial B_{\mathrm{H}}^2} & \dfrac{\partial^2 f}{\partial B_{\mathrm{H}} \partial B_{\mathrm{L}}} \\ \dfrac{\partial^2 f}{\partial I \partial B_{\mathrm{L}}} & \dfrac{\partial^2 f}{\partial B_{\mathrm{H}} \partial B_{\mathrm{L}}} & \dfrac{\partial^2 f}{\partial B_{\mathrm{L}}^2} \end{bmatrix}\begin{bmatrix} 1 \\ g'_{\mathrm{H}} \\ g'_{\mathrm{L}} \end{bmatrix} + \left[\frac{\partial f}{\partial B_{\mathrm{H}}}, \ \frac{\partial f}{\partial B_{\mathrm{L}}}\right]\begin{bmatrix} g''_{\mathrm{H}} \\ g''_{\mathrm{L}} \end{bmatrix} \end{aligned}$$

which matches Equation (25).

COROLLARY 5.3 *Let $X(t)$ be a non-trivial solution of system* (1)–(4). *Also assume $R_0 > 1$. If $\lim_{t\to\infty} X(t) = X_\infty$ exists, then $X_\infty = X^*$, the positive endemic equilibrium.*

Proof The assumption $\lim_{t\to\infty} X(t) = X_\infty$ exists which implies that $\lim_{t\to\infty} X'(t) = 0$, which can be shown easily by the standard $\varepsilon - \delta$ argument.

Taking the limit $t \to \infty$ in system (1)–(4), we obtain $F(X_\infty) = 0$. Hence, X_∞ must be an equilibrium point of the system. Since $R_0 > 1$, the DFE is unstable (see Theorem 3.2) and the system has a unique non-trivial equilibrium X^* based on Theorem 5.1. Therefore, $X_\infty = X^*$. ■

Remark 5.4 Note the assumption that $\lim\limits_{t\to\infty} X(t) = X_\infty$ exists in Corollary 5.3. This condition can be (largely) leveraged through our stability analysis in what follows.

6. Stability of the endemic equilibrium

6.1. *Local stability*

Now that we have established the existence of the unique positive endemic equilibrium X^*, we proceed to show that X^* is locally asymptotically stable.

The Jacobian matrix of the model system (1)–(4) is

$$J_B = \begin{bmatrix} -b - f(I,B) & -S\dfrac{\partial f}{\partial I}(I,B) & 0 & -S\dfrac{\partial f}{\partial B}(I,B) \\ f(I,B) & S\dfrac{\partial f}{\partial I}(I,B) - (\gamma + b) & 0 & S\dfrac{\partial f}{\partial B}(I,B) \\ 0 & \gamma & -b & 0 \\ 0 & \dfrac{\partial h}{\partial I}(I,B) & 0 & \dfrac{\partial h}{\partial B}(I,B) \end{bmatrix}. \tag{26}$$

At the endemic equilibrium $X^* = (S^*, I^*, R^*, B^*)$, the components satisfy

$$I^* = \frac{1}{(\gamma + b)} \frac{bNf(I^*, B^*)}{b + f(I^*, B^*)}, \tag{27}$$

$$S^* = \frac{bN}{b + f(I^*, B^*)}, \tag{28}$$

$$R^* = \frac{\gamma}{b} I^*, \tag{29}$$

$$0 = h(I^*, B^*). \tag{30}$$

For the convenience of algebraic manipulation, we denote

$$F = f(I^*, B^*), \quad E = \frac{\partial f}{\partial I}(I^*, B^*), \quad P = \frac{\partial f}{\partial B}(I^*, B^*), \quad Q = \frac{\partial h}{\partial B}(I^*, B^*), \quad T = \frac{\partial h}{\partial I}(I^*, B^*).$$

From the assumptions (b) and (c), $F \geq 0, E \geq 0, P \geq 0, T \geq 0$, whereas $Q \leq 0$. Evaluated at X^*, the Jacobian matrix (26) becomes

$$J_B^* = \begin{bmatrix} -F - b & -S^*E & 0 & -S^*P \\ F & S^*E - (\gamma + b) & 0 & S^*P \\ 0 & \gamma & -b & 0 \\ 0 & T & 0 & Q \end{bmatrix}.$$

The characteristic polynomial of J_B^* is

$$\begin{aligned} \text{Det}\,(\lambda I - J_B^*) = (\lambda + b)[&(\lambda + b)(\lambda - S^*E + \gamma + b)(\lambda - Q) \\ &+ F(\lambda + \gamma + b)(\lambda - Q) - (\lambda + b)S^*PT]. \end{aligned}$$

The equilibrium X^* is locally asymptotically stable if and only if all roots of the above polynomial have negative real parts. Obviously, $\lambda = -b$ is a negative root. To investigate the other three roots,

we expand the expression in the square brackets to obtain a cubic equation

$$a_0\lambda^3 + a_1\lambda^2 + a_2\lambda^1 + a_3 = 0, \tag{31}$$

with

$$a_0 = 1, \tag{32}$$
$$a_1 = F - Q + 2b + \gamma - ES^*, \tag{33}$$
$$a_2 = b^2 - FQ + Fb + F\gamma - 2Qb - Q\gamma + b\gamma + EQS^* - PS^*T - ES^*b, \tag{34}$$
$$a_3 = -Qb^2 - FQb - FQ\gamma - Qb\gamma + EQS^*b - PS^*Tb. \tag{35}$$

To ensure that all roots of Equation (31) have negative real parts, the Routh–Hurwitz stability criterion [23] requires

$$a_1 > 0, \quad a_2 > 0, \quad a_3 > 0, \quad a_1a_2 > a_0a_3. \tag{36}$$

We shall prove all the four inequalities in Equation (36). To that end we first establish the following lemma:

LEMMA 6.1 *At the endemic equilibrium X^*, we have*

$$b + \gamma - ES^* \geq 0 \tag{37}$$
$$-Q(b + \gamma) \geq PTS^* - EQS^*. \tag{38}$$

Proof Based on our assumption (d), we have known that the inequality (16) holds at any given point $(I_0, B_0) \geq 0$. In particular, if we set $(I_0, B_0) = (I^*, B^*)$, i.e., the positive endemic equilibrium, we obtain

$$f(I,B) \leq f(I^*,B^*) + \frac{\partial f}{\partial I}(I^*,B^*)(I - I^*) + \frac{\partial f}{\partial B}(I^*,B^*)(B - B^*), \tag{39}$$

which holds for all $(I, B) \geq 0$. Substitute $B = B^*$, $I = 0$ and Equation (39) becomes

$$0 \leq f(0,B^*) \leq f(I^*,B^*) - \frac{\partial f}{\partial I}(I^*,B^*)I^*. \tag{40}$$

Using Equations (27) and (28) and inequality (40), we obtain

$$\begin{aligned} b + \gamma - ES^* &= (b + \gamma) - \frac{\partial f}{\partial I}(I^*,B^*)S^* \\ &= \frac{bNf(I^*,B^*)}{[b + f(I^*,B^*)]I^*} - \frac{\partial f}{\partial I}(I^*,B^*)\frac{bN}{b + f(I^*,B^*)} \\ &= \frac{bN}{[b + f(I^*,B^*)]I^*}\left[f(I^*,B^*) - \frac{\partial f}{\partial I}(I^*,B^*)I^*\right] \\ &\geq 0, \end{aligned} \tag{41}$$

which establishes the result in Equation (37).

Next, based on the assumption (e), the function $h(I,B)$ is concave at the point (I^*,B^*). Thus

$$h(I,B) \le h(I^*,B^*) + \frac{\partial h}{\partial I}(I^*,B^*)(I-I^*) + \frac{\partial h}{\partial B}(I^*,B^*)(B-B^*). \tag{42}$$

Note that $h(I^*,B^*)=0$, $h(0,0)=0$. Substitute $I=B=0$ into Equation (42) to obtain

$$\frac{\partial h}{\partial I}(I^*,B^*)I^* + \frac{\partial h}{\partial B}(I^*,B^*)B^* \le 0. \tag{43}$$

Since $(\partial h/\partial B)(I^*,B^*) \le 0$ due to the assumption (c), the inequality (43) yields

$$\overline{B} \triangleq B^* + \frac{(\partial h/\partial I)(I^*,B^*)}{(\partial h/\partial B)(I^*,B^*)} I^* \ge 0. \tag{44}$$

Now, substitute the point $(I,B)=(0,\overline{B})$, which is in the biologically feasible domain of our model, into the inequality (39) to obtain

$$0 \le f(0,\overline{B}) \le f(I^*,B^*) - \frac{\partial f}{\partial I}(I^*,B^*)I^* + \frac{\partial f}{\partial B}(I^*,B^*)\frac{(\partial h/\partial I)(I^*,B^*)}{(\partial h/\partial B)(I^*,B^*)}I^*. \tag{45}$$

Combining the inequality (45) and the facts: $S^*f(I^*,B^*) = (\gamma+b)I^*$, $(\partial h/\partial B)(I^*,B^*) \le 0$, we obtain

$$\begin{aligned} -Q(b+\gamma) &= -\frac{\partial h}{\partial B}(I^*,B^*)(b+\gamma) \\ &\ge \frac{\partial f}{\partial B}(I^*,B^*)\frac{\partial h}{\partial I}(I^*,B^*)S^* - \frac{\partial f}{\partial I}(I^*,B^*)\frac{\partial h}{\partial B}(I^*,B^*)S^* \\ &= PTS^* - EQS^*, \end{aligned} \tag{46}$$

which establishes the result in Equation (38). ■

Based on Lemma 6.1, we are now ready to proceed to Equation (36).

LEMMA 6.2 *At the endemic equilibrium X^*, all the four inequalities in Equation* (36) *hold.*

Proof First, using the inequality (37), we obtain

$$\begin{aligned} a_1 &= F - Q + 2b + \gamma - ES^* \\ &= f(I^*,B^*) - \frac{\partial h}{\partial B}(I^*,B^*) + 2b + \gamma - \frac{\partial f}{\partial I}(I^*,B^*)S^* \\ &> (b+\gamma) - \frac{\partial f}{\partial I}(I^*,B^*)S^* \\ &> 0. \end{aligned} \tag{47}$$

Next, using both results in Equations (37) and (38), we obtain

$$\begin{aligned} a_2 &= b^2 - FQ + Fb + F\gamma - 2Qb - Q\gamma + b\gamma + EQS^* - PS^*T - ES^*b \\ &= b(b+\gamma-ES^*) + (-Qb - Q\gamma - PS^*T + EQS^*) + (Fb + F\gamma - FQ - Qb) \\ &> 0. \end{aligned} \tag{48}$$

Similarly, we have

$$\begin{aligned} a_3 &= -Qb^2 - FQb - FQ\gamma - Qb\gamma + EQS^*b - PS^*Tb \\ &= b(-Qb - Q\gamma + EQS^* - PS^*T) + (-FQb - FQ\gamma) \\ &> 0. \end{aligned} \tag{49}$$

Finally, note that $a_1 = F - Q + 2b + \gamma - ES^* > -Q > 0$ and that

$$(-Q)a_2 - a_0a_3 = (Q^2b + Q^2\gamma - EQ^2S^* + PTQS^*) + (FQ^2 + Q^2b + PS^*Tb) > 0. \tag{50}$$

It is thus clear to see $a_1a_2 > a_0a_3$ holds. ■

Therefore, based on the Routh–Hurwitz stability criterion, we have established the following result:

THEOREM 6.3 *When $R_0 > 1$, the endemic equilibrium of system* (1)–(4) *is locally asymptotically stable.*

6.2. *Linear global stability*

To prove the global asymptotic stability of the endemic equilibrium, the key is to show the non-existence of periodic orbits. This is generally difficult for high-dimensional model systems, as the classical Poincaré–Bendixson framework [13] is no longer valid in high dimensions. For some special cases, however, our model (1)–(4) can be reduced to a two-dimensional autonomous system in S and I, and classical dynamical system theory can be applied. Below, we present two simplified cases with linear and bilinear incidence rates, respectively.

First, we assume that the incidence $f(I, B) = C$, where $C > 0$ is a constant. Our model is then reduced to a two-dimensional linear system

$$\frac{\mathrm{d}S}{\mathrm{d}t} = bN - (C + b)S, \tag{51}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = CS - (\gamma + b)I. \tag{52}$$

In such a case, our assumption (a) is not valid so that there is no DFE. The epidemiological implication is that the pathogen concentrations and/or the infected numbers are at such a high level that the infection is certain to those exposed. There is a unique positive endemic equilibrium of the system (51) and (52):

$$S^* = \frac{bN}{C + b} \quad \text{and} \quad I^* = \frac{bCN}{(C + b)(\gamma + b)}. \tag{53}$$

Since system (51) and (52) is linear, its exact solution can be easily found as

$$S(t) = \frac{bN}{C + b} + \left(S(0) - \frac{bN}{C + b}\right) \mathrm{e}^{-(C+b)t}$$

$$I(t) = \frac{bCN}{(C + b)(\gamma + b)} + k_1 \, \mathrm{e}^{-(C+b)t} + k_2 \, \mathrm{e}^{-(\gamma+b)t}$$

with

$$k_1 = \frac{C}{\gamma - C}\left(S(0) - \frac{bN}{C + b}\right), \quad k_2 = I(0) - \frac{bCN}{(C + b)(\gamma + b)} - \frac{C}{\gamma - C}\left(S(0) - \frac{bN}{C + b}\right).$$

It is clear to see that $S(t) \to S^*$ and $I \to I^*$ as $t \to \infty$, regardless of the initial values of S and I. Hence, the endemic equilibrium (S^*, I^*) is globally asymptotically stable.

In the second case, we assume $f(I, B) = CI$, where $C > 0$ is a constant. The original system (1)–(4) is then reduced to

$$\frac{\mathrm{d}S}{\mathrm{d}t} = F_1(S, I) = b(N - S) - CIS, \tag{54}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = F_2(S, I) = CIS - (\gamma + b)I, \tag{55}$$

which represents a regular SIR model with a normal bilinear incidence. The endemic equilibrium of this simplified model is

$$(S^*, I^*) = \left(\frac{\gamma + b}{C}, \frac{bN}{\gamma + b} - \frac{b}{C}\right). \tag{56}$$

Note that $R_0 = (N/(\gamma + b))C$ (see Equation (11)). Hence $I^* = (bN/(\gamma + b)) - b/C > 0$ due to $R_0 > 1$. Applying the Dulac's criterion [13,15,27] and introducing an auxiliary function $P(S, I) = 1/I$, we obtain

$$\frac{\partial}{\partial S}(PF_1) + \frac{\partial}{\partial I}(PF_2) = -\left(C + \frac{b}{I}\right) < 0,$$

which holds everywhere in the region

$$D = \{(S, I) \mid S > 0, I > 0, S + I < N\}. \tag{57}$$

Hence, there is no periodic solution in D and the endemic equilibrium (S^*, I^*) is globally asymptotically stable.

When the incidence rates become nonlinear, system (1)–(4) generally has dimensions higher than two. For several special types of cholera models with nonlinear incidence rates, the global endemic stability was analysed in detail in a recent work [50]. These linear and nonlinear results motivate the speculation that the endemic equilibrium X^* of system (1)–(4) is globally asymptotically stable in general (provided $R_0 > 1$). The proof for the general model has not been resolved yet, and we plan to explore this topic in our future research.

6.3. *Bifurcation diagram*

Our stability analysis of the DFE and the endemic equilibrium shows a supercritical bifurcation with respect to the parameter R_0. The results are summarized by the following theorem:

THEOREM 6.4 *Under the assumptions* (a)–(f), *model system* (1)–(4) *has a forward transcritical bifurcation at* $R_0 = 1$.

Remark 6.5 Theorem 6.4 states that for biologically feasible incidence and pathogen functions, our cholera model does not exhibit backward (or subcritical) bifurcation [4,7,44,47] and the endemic level is continuously depending on R_0. A value of R_0 slightly above 1 will, regarding long-term dynamics, only lead to a low endemic state. This has important implications for the prevention and intervention strategies for cholera, as reducing, and keeping, R_0 below 1 would be sufficient to eradicate the disease in the long run.

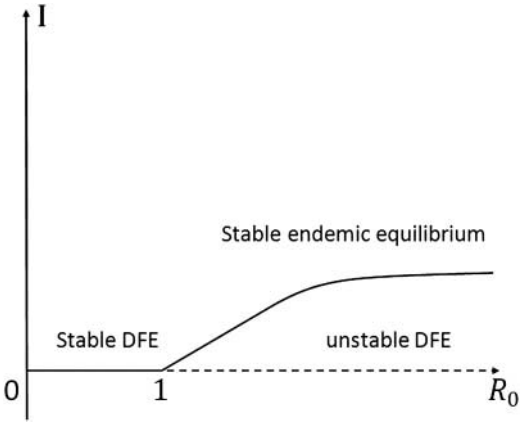

Figure 2. The bifurcation diagram of I vs. R_0 which shows a transcritical bifurcation at $R_0 = 1$.

A bifurcation diagram for I vs. R_0 is sketched in Figure 2. At the DFE, $I = 0$ is stable for $R_0 < 1$ and unstable for $R_0 > 1$. At the endemic equilibrium,

$$I = \tilde{H}(I) = \tilde{H}(0) + \tilde{H}'(0)I + Q(I), \tag{58}$$

where $\tilde{H}(I)$ is defined in Equation (20) and where

$$Q(I) = \sum_{m=2}^{\infty} \frac{\tilde{H}^m(0)}{m!} I^m. \tag{59}$$

Since $\tilde{H}(0) = 0$, $\tilde{H}'(0) = R_0$, we obtain

$$I = R_0 I + Q(I), \quad \text{or} \quad R_0 = 1 - \frac{Q(I)}{I}. \tag{60}$$

Based on Equation (60), when I is small, $R_0 \approx 1 - (\tilde{H}''(0)/2)I$ (notice $\tilde{H}''(0) \leq 0$), which is approximately a straight line passing the bifurcation point $(R_0, I) = (1, 0)$. When $I \to \infty$, $(\mathrm{d}R_0/\mathrm{d}I) \to \infty$, so that the endemic equilibrium curve becomes more and more horizontal.

7. Examples

Our generalized models (1)–(4) can unify many existing cholera models [5,11,12,18,40,43,51], so that these different models can be studied and applied through a single unified framework. Below, we briefly discuss three representative models.

7.1. *The model of Codeço*

This model is given by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = n(H - S) - a\frac{B}{\kappa + B}S, \tag{61}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = a\frac{B}{\kappa + B}S - rI, \tag{62}$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = eI - (mb - nb)B, \tag{63}$$

where H stands for the total population, and $mb - nb > 0$ represents the net death rate of vibrios. In this model, the incidence is $f(I,B) = a(B/(\kappa + B))$ and the pathogen function is $h(I,B) = eI - (mb - nb)B$. Only the environment-to-human transmission mode is considered in this formulation.

It can be easily verified that the assumptions (a)–(f) all hold for system (61)–(63). Hence, all the analytical results presented so far can be applied to this model. In particular, the basic reproduction number R_0 is determined by Equation (11):

$$R_0 = \frac{N}{\gamma + b}\left[\frac{\partial f}{\partial I}(0,0) + \frac{\partial f}{\partial B}(0,0)g'(0)\right] = \frac{Nae}{\kappa(\gamma + b)(mb - nb)},$$

which agrees with the result obtained earlier [5].

7.2. *The model of Mukandavire et al.*

This model is based on the following differential equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \mu N - \beta_{\mathrm{e}} S\frac{B}{\kappa + B} - \beta_{\mathrm{h}} SI - \mu S, \tag{64}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta_{\mathrm{e}} S\frac{B}{\kappa + B} + \beta_{\mathrm{h}} SI - (\gamma + \mu)I, \tag{65}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \mu R, \tag{66}$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \xi I - \delta B. \tag{67}$$

The incidence is $f(I,B) = \beta_{\mathrm{e}}(B/(\kappa + B)) + \beta_{\mathrm{h}} I$, and $h(I,B) = \xi I - \delta B$. Both environment-to-human and human-to-human transmission modes are included in this model. It is straightforward to verify the assumptions (a)–(f); in particular,

$$D^2 f = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{-2\kappa\beta_e}{(\kappa + B)^3} \end{bmatrix} \quad \text{and} \quad D^2 h = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

are both negative semidefinite for all $I, B \geq 0$. Based on Equation (11), the basic reproduction number is

$$R_0 = \frac{N}{\gamma + b}\left[\beta_{\mathrm{h}} + \frac{\beta_{\mathrm{e}}}{\kappa}\frac{\xi}{\delta}\right] = \frac{N}{\delta\kappa(\gamma + b)}(\kappa\delta\beta_{\mathrm{h}} + \xi\beta_{\mathrm{e}}).$$

The same result was obtained in the work of Mukandavire *et al*. [40].

7.3. *The model of Hartley et al.*

The model explicitly incorporates both the HI and LI states of *V. cholerae*, and takes the form

$$\frac{\mathrm{d}S}{\mathrm{d}t} = bN - \beta_{\mathrm{L}} S\frac{B_{\mathrm{L}}}{\kappa_{\mathrm{L}} + B_{\mathrm{L}}} - \beta_{\mathrm{H}} S\frac{B_{\mathrm{H}}}{\kappa_{\mathrm{H}} + B_{\mathrm{H}}} - bS, \tag{68}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta_{\mathrm{L}} S\frac{B_{\mathrm{L}}}{\kappa_{\mathrm{L}} + B_{\mathrm{L}}} + \beta_{\mathrm{H}} S\frac{B_{\mathrm{H}}}{\kappa_{\mathrm{H}} + B_{\mathrm{H}}} - (\gamma + b)I, \tag{69}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - bR, \tag{70}$$

$$\frac{\mathrm{d}B_{\mathrm{H}}}{\mathrm{d}t} = \xi I - \chi B_{\mathrm{H}}, \tag{71}$$

$$\frac{\mathrm{d}B_{\mathrm{L}}}{\mathrm{d}t} = \chi B_{\mathrm{H}} - \delta_{\mathrm{L}} B_{\mathrm{L}}. \tag{72}$$

Here

$$B = [B_{\mathrm{H}}, B_{\mathrm{L}}], \quad f(I,B) = \beta_{\mathrm{L}} \frac{B_{\mathrm{L}}}{\kappa_{\mathrm{L}} + B_{\mathrm{L}}} + \beta_{\mathrm{H}} S \frac{B_{\mathrm{H}}}{\kappa_{\mathrm{H}} + B_{\mathrm{H}}}, \quad \text{and} \quad h(I,B) = \begin{bmatrix} \xi I - \chi B_{\mathrm{H}} \\ \chi B_{\mathrm{H}} - \delta_{\mathrm{L}} B_{\mathrm{L}} \end{bmatrix}.$$

The assumptions (a)–(f) can be similarly verified. For instance,

$$\frac{\partial f}{\partial B} = \left[\frac{\beta_{\mathrm{H}} \kappa_{\mathrm{H}}}{(\kappa_{\mathrm{H}} + B_{\mathrm{H}})^2}, \frac{\beta_{\mathrm{L}} \kappa_{\mathrm{L}}}{(\kappa_{\mathrm{L}} + B_{\mathrm{L}})^2} \right] > 0, \quad \frac{\partial h}{\partial B} = \begin{bmatrix} -\chi & 0 \\ \chi & -\delta_{\mathrm{L}} \end{bmatrix} < 0 \quad \text{(negative definite)}.$$

Meanwhile,

$$D^2 f = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \dfrac{-2\beta_{\mathrm{H}} \kappa_{\mathrm{H}}}{(\kappa_{\mathrm{H}} + B_{\mathrm{H}})^3} & 0 \\ 0 & 0 & \dfrac{-2\beta_{\mathrm{L}} \kappa_{\mathrm{L}}}{(\kappa_{\mathrm{L}} + B_{\mathrm{L}})^3} \end{bmatrix}$$

and $D^2 h = 0$ are both negative semidefinite for all $I \geq 0, B \geq 0$. The basic reproduction number for this model is

$$\begin{aligned} R_0 &= \frac{N}{\gamma + b} \left[\frac{\partial f}{\partial I}(0,0) + \frac{\partial f}{\partial B}(0,0) g'(0) \right] \\ &= \frac{N}{\gamma + b} \left[0 + \left(\frac{\beta_{\mathrm{H}}}{\kappa_{\mathrm{H}}}, \frac{\beta_{\mathrm{L}}}{\kappa_{\mathrm{L}}} \right) \begin{pmatrix} \xi/\chi \\ \xi/\delta_{\mathrm{L}} \end{pmatrix} \right] \\ &= \frac{N\xi}{\gamma + b} \left(\frac{\beta_{\mathrm{H}}}{\kappa_{\mathrm{H}} \chi} + \frac{\beta_{\mathrm{L}}}{\kappa_{\mathrm{L}} \delta_{\mathrm{L}}} \right) \end{aligned}$$

which exactly matches the result given by Hartley *et al.* [12].

A quantitative comparison of these three models is made by applying each of them to study the epidemic and endemic cholera dynamics in a hypothetical community with a total population of $N = 10,000$. The initial condition is set as $I(0) = 1,\ S(0) = N - 1,\ R = B = 0$; i.e., one infective initially enters the wholly susceptible community. The parameter values are based on the cholera data published on Zimbabwe [35,40][1,2,3]. Figure 3 shows the simulation results of the infection curves for the three models. The first peak in each curve represents the cholera outbreak triggered by the initial infection. Among the three, the model of Hartley *et al.* shows the highest infection number because of its explicit incorporation of the HI state of the vibrios, whereas the model of Codeço exhibits the lowest epidemic value as it considered only in the environment-to-human transmission pathway with an LI state of the pathogen. After the first cholera peak, all the three infection curves decline and show several outbreaks with decaying magnitudes, before they finally rest at their endemic equilibria. The model of Mukandavire *et al.* exhibits a few more epidemic oscillations than the other two models because of its explicit inclusion of both the environment-to-human and human-to-human transmission modes which lead to longer epidemic dynamics. We found the endemic infection equilibria are $I^* \doteq 0.88, 0.75, 0.92$ for the model of Codeço, that of Mukandavire *et al.*, and that of Hartley *et al.*, respectively, implying relatively low endemicity. If we scale up these numbers using the realistic population size in Zimbabwe (about

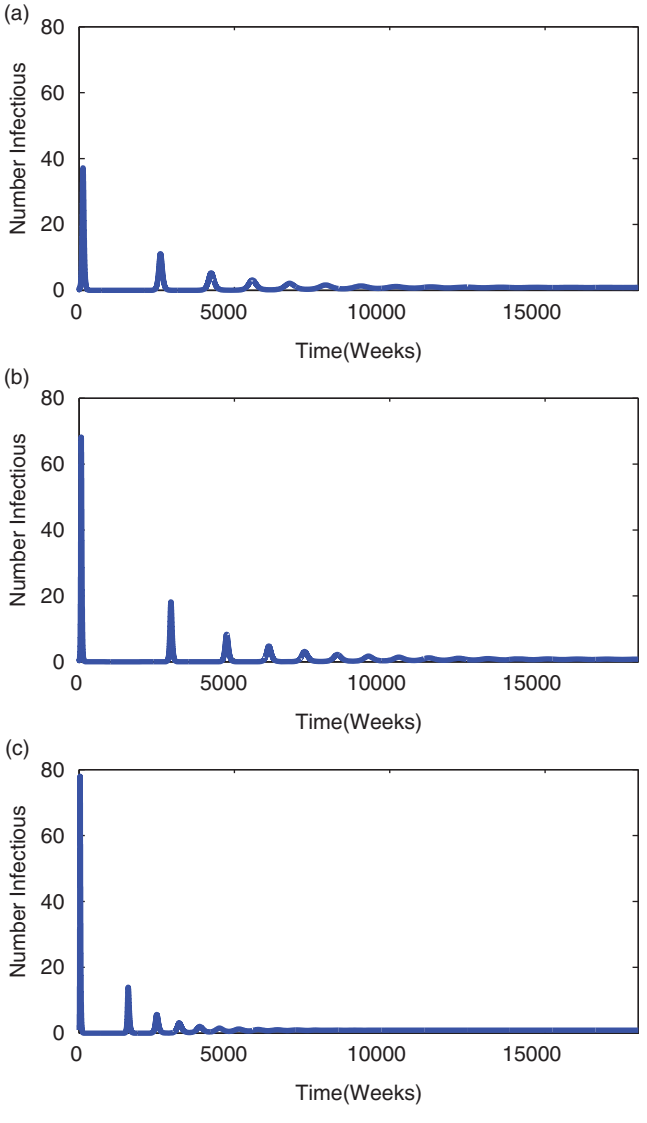

Figure 3. Simulation results for a hypothetical community with a total population of $N = 10,000$, using three different cholera models: (a) the model of Codeço [5]; (b) the model of Mukandavire *et al.* [40]; and (c) the model of Hartley *et al.* [12]. The initial condition is $I(0) = 1, S(0) = N - 1, R = B = 0$. After the major cholera outbreak (i.e., the first peak) caused by the initial infection, each infection curve exhibits several small-scale epidemic oscillations and finally converges to the endemic equilibrium over time. The endemic values are $I^* \doteq 0.88, 0.75, 0.92$ for the three models, respectively.

13.3 million in 2010[1]), we obtain that the endemic infection numbers would be about 1164, 992, and 1217, respectively, based on these model predictions.

8. Discussion

We have presented a generalized mathematical cholera model and conducted an analysis for the epidemic and endemic dynamics. By introducing general incidence and pathogen functions, our model can unify cholera studies into a single framework of modelling, simulation, and analysis. Meanwhile, new model development and analysis can be possibly made in the same framework as well.

For simple SI epidemiological models with constant population, Lajmanovich and Yorke [25] proved that the DFE is globally asymptotically stable when $R_0 < 1$, and there is a unique endemic equilibrium which is globally asymptotically stable when $R_0 > 1$. Such a prototypical R_0 threshold behaviour has since been extensively studied for many infectious diseases with various SI, SIR, SEI and SEIR models (see [4,16,28,30,31,37,46,55], among others).

The complication of cholera dynamics lies in the coupling between human hosts and environmental components which leads to a combined human–environment epidemiological model. Nonetheless, the stability analysis in the present paper shows that under biologically feasible conditions, a regular (i.e., forward) transcritical bifurcation occurs at $R_0 = 1$. Specifically, we have established that for $R_0 < 1$, there is a unique DFE which is both locally and globally asymptotically stable; this equilibrium becomes unstable when $R_0 > 1$. Meanwhile, there is a unique positive endemic equilibrium which is locally asymptotically stable when $R_0 > 1$. This is important for the prevention and intervention strategies against cholera outbreaks, as reducing and keeping R_0 below 1 would be sufficient to eradicate the disease in the long term.

It remains to show the global asymptotic stability for the endemic equilibrium of our unified cholera model. This is generally difficult for high-dimensional nonlinear systems, such as ours in (1)–(4). Lajmanovich and Yorke [25] carefully constructed two Lyapunov functions [20] to prove the global asymptotic stability of the DFE and the endemic equilibrium. However, the fact that there is no systematic way to find Lyapunov functions poses a challenge to the application of this approach to general nonlinear systems. Quite a few efforts [9,10,27,36,45,47,48] have been devoted to extend the classical Poincaré–Bendixson framework [13] to high-dimensional systems, using theory of monotone flows, competitive systems, and Lipschitz manifolds, etc., though such extensions are in general highly non-trivial. Finally, in a series of papers, Li and co-workers [28,29,31,32] analysed the global stability of high-dimensional endemic equilibria based on the theory of monotone dynamical systems and geometric approaches. Similar work of epidemiological global dynamics was also conducted in earlier studies [30,39,55]. Some of these approaches were applied in a recent work [50] to analyse several special types of cholera models. These studies provide useful directions for our future work on the global endemic stability of the most general cholera model.

In addition, the cholera model and analysis proposed in this paper can be extended in a number of ways. For example, climatic impacts (such as rainfall, monsoon, flood, drought, and water temperature) on cholera epidemics [22,34] can be studied by incorporating seasonally variational factors into the incidence function f. Meanwhile, several recent studies [8,18,41] have suggested that cholera dynamics is closely related to the prevalence of bacteriophages in the environment; the effects of those vibriophages on cholera can be easily added to the environmental function h in our model. Furthermore, prevention and intervention strategies, such as vaccination, water sanitation, hydration therapy, and antibiotic treatment, can be naturally represented by modifying the two functions f and h in our model, so as to seek possible optimal control strategies [26] against cholera outbreaks.

Acknowledgement

This work was partially supported by the National Science Foundation under grant no. DMS-0813691. The authors would also like to thank the anonymous reviewers for their valuable comments.

Notes

1. World Health Organization web page: www.who.org.
2. Center for Disease Control and Prevention web page: www.cdc.gov.
3. Wikipedia: en.wikipedia.org.

References

[1] A. Alam, R.C. LaRocque, J.B. Harris, C. Vanderspurt, E.T. Ryan, F. Qadri, and S.B. Calderwood, *Hyperinfectivity of human-passaged Vibrio cholerae can be modeled by growth in the infant mouse*, Infect. Immun. 73 (2005), pp. 6674–6679.

[2] V. Capasso and S.L. Paveri-Fontana, *A mathematical model for the 1973 cholera epidemic in the European Mediterranean region*, Rev. dépidémoligié et de santé Publiqué 27 (1979), pp. 121–132.

[3] C. Castillo-Chavez, Z. Feng, and W. Huang, *On the Computation of R_0 and its Role on Global Stability*, Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction, IMA Vol. 125, Springer-Verlag, Berlin, 2002.

[4] N. Chitnis, J.M. Cushing, and J.M. Hyman, *Bifurcation analysis of a mathematical model for malaria transmission*, SIAM J. Appl. Math. 67 (2006), pp. 24–45.

[5] C.T. Codeço, *Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir*, BMC Infect. Dis. 1 (2001), p. 1.

[6] K. Dietz, *The estimation of the basic reproduction number for infections diseases*, Stat. Methods Med. Res. 2 (1993), pp. 23–41.

[7] J. Dushoff, W. Huang, and C. Castillo-Chavez, *Backwards bifurcation and catastrophe in simple models of fatal diseases*, J. Math. Biol. 36 (1998), pp. 227–248.

[8] S.M. Faruque, I.B. Naser, M.J. Islam, A.S.G. Faruque, A.N. Ghosh, G.B. Nair, D.A. Sack, and J.J. Mekalanos, *Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages*, Proc. Natl Acad. Sci. 102 (2005), pp. 1702–1707.

[9] M. Feckan, *A generalization of Bendixson's criterion*, Proc. Am. Math. Soc. 129 (2001), pp. 3395–3399.

[10] M. Feckan, *Criteria on the nonexistence of invariant Lipschitz submanifolds for dynamical systems*, J. Diff. Eq. 174 (2001), pp. 392–419.

[11] M. Ghosh, P. Chandra, P. Sinha, and J. B. Shukla, *Modeling the spread of carrier-dependent infectious diseases with environmental effect*, Appl. Math. Comput. 152 (2004), pp. 385–402.

[12] D.M. Hartley, J.G. Morris, and D.L. Smith, *Hyperinfectivity: a critical element in the ability of* V. cholerae *to cause epidemics?* PLoS Med. 3 (2006), pp. 0063–0069.

[13] P. Hartman, *Ordinary Differential Equations*, John Wiley, New York, 1980.

[14] T.R. Hendrix, *The pathophysiology of cholera*, Bull. NY Acad. Med. 47 (1971), pp. 1169–1180.

[15] H.W. Hethcote, *Qualitative analysis of communicable disease models*, Math. Biosci. 28 (1976), pp. 335–356.

[16] H.W. Hethcote, *The mathematics of infectious diseases*, SIAM Rev. 42 (2000), pp. 599–653.

[17] H.W. Hethcote and S.A. Levin, *Periodicity in epidemiological models*, in *Applied Mathematical Ecology*, Levin, Hallam, Gross, eds., Springer-Verlag, Berlin, 1989.

[18] M.A. Jensen, S.M. Faruque, J.J. Mekalanos, and B.R. Levin, *Modeling the role of bacteriophage in the control of cholera outbreaks*, Proc. Natl Acad. Sci. 103 (2006), pp. 4652–4657.

[19] J.B. Kaper, J.G. Morris, and M.M. Levine, *Cholera*, Clin. Microbiol. Rev. 8 (1995), pp. 48–86.

[20] H.K. Khalil, *Nonlinear Systems*, Prentice Hall, Upper Saddle River, NJ, 1996.

[21] A.A. King, E.L. Lonides, M. Pascual, and M.J. Bouma, *Inapparent infections and cholera dynamics*, Nature 454 (2008), pp. 877–881.

[22] K. Koelle, X. Rodo, M. Pascual, Md. Yunus, and G. Mostafa, *Refractory periods and climate forcing in cholera dynamics*, Nature 436 (2005), pp. 696–700.

[23] G.A. Korn and T.M. Korn, *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for References and Review*, Dover, Mineola, NY, 2000.

[24] A. Korobeinikov and P.K. Maini, *Non-linear incidence and stability of infectious disease models*, Math. Med. Biol. 22 (2005), pp. 113–128.

[25] A. Lajmanovich and J. Yorke, *A deterministic model for gonorrhea in a nonhomogeneous population*, Math. Biosci. 28 (1976), pp. 221–236.

[26] S. Lenhart and J. Workman, *Optimal Control Applied to Biological Models*, Chapman Hall/CRC, London, 2007.

[27] B. Li, *Periodic orbits of autonomous ordinary differential equations: theory and applications*, Nonlinear Anal.: Theory, Methods Appl. 5 (1981), pp. 931–958.

[28] M.Y. Li and J.S. Muldowney, *Global stability for the SEIR model in epidemiology*, Math. Biosci. 125 (1995), pp. 155–164.

[29] M.Y. Li and L. Wang, *Global stability in some SEIR epidemic models*, IMA Vol. Math. Appl. 126 (2002), pp. 295–311.

[30] G. Li and J. Zhen, *Global stability of an SEI epidemic model with general contact rate*, Chaos, Solitons Fractals 23 (2005), pp. 997–1004.

[31] M.Y. Li, J.R. Graef, L. Wang, and J. Karsai, *Global dynamics of a SEIR model with varying total population size*, Math. Biosci. 160 (1999), pp. 191–213.

[32] M.Y. Li, J.S. Muldowney, and P.V.D. Driessche, *Global stability of SEIRS models in epidemiology*, Can. Appl. Math. Q. 7 (1999), pp. 409–425.

[33] S. Liao and J. Wang, *Stability analysis and application of a mathematical cholera model*, Math. Biosci. Eng. 8 (2011), pp. 733–752.

[34] B. Lobitz, L. Beck, A. Huq, B. Wood, G. Fuchs, A.S.G. Faruque, and R. Colwell, *Climate and infectious disease: use of remote sensing for detection of* Vibrio cholerae *by indirect measurement*, Proc. Natl Acad. Soc. USA 97 (2000), pp. 1438–1443.

[35] P.R. Mason, *Zimbabwe experiences the worst epidemic of cholera in Africa*, J. Infect. Dev. Countries 3 (2009), pp. 148–151.
[36] V.S. Medvedev, *Sufficient conditions for dynamical systems on manifolds to have no integral cycles*, Diff. Equations 6 (1970), pp. 454–466.
[37] J. Mena-Lorca and H.W. Hethcote, *Dynamic models of infectious diseases as regulator of population sizes*, J. Math. Biol. 30 (1992), pp. 693–716.
[38] D.S. Merrell, S.M. Butler, and F. Qadri *et al.*, *Host-induced epidemic spread of the cholera bacterium*, Nature 417 (2002), pp. 642–645.
[39] S.M. Moghadas and A.B. Gumel, *Global stability of a two-stage epidemic model with generalized non-linear incidence*, Math. Comput. Simul. 60 (2002), pp. 107–118.
[40] 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*, Proc. Natl Acad. Sci. 108 (2011), pp. 8767–8772.
[41] E.J. Nelson, J.B. Harris, J.G. Morris, S.B. Calderwood, and A. Camilli, *Cholera transmission: the host, pathogen and bacteriophage dynamics*, Nat. Rev.: Microbiology 7 (2009), pp. 693–702.
[42] M. Pascual, M. Bouma, and A. Dobson, *Cholera and climate: revisiting the quantitative evidence*, Microbes Infections 4 (2002), pp. 237–245.
[43] E. Pourabbas, A. d'Onofrio, and M. Rafanelli, *A method to estimate the incidence of communicable diseases under seasonal fluctuations with application to cholera*, Appl. Math. Comput. 118 (2001), pp. 161–174.
[44] T.C. Reluga, J. Medlock, and A.S. Perelson, *Backward bifurcation and multiple equilibria in epidemic models with structured immunity*, J. Theoret. Biol. 252 (2008), pp. 155–165.
[45] L.A. Sanchez, *Existence of periodic orbits for high-dimensional autonomous systems*, J. Math. Anal. Appl. 363 (2010), pp. 409–418.
[46] C.P. Simon and J.A. Jacquez, *Reproduction numbers and the stability of equilibria of SI models for heterogeneous populations*, SIAM J. Appl. Math. 52 (1992), pp. 541–576.
[47] B.H. Singer and D. Kirschner, *Influence of backward bifurcation on interpretation of R_0 in a model of epidemic tuberculosis with reinfection*, Math. Biosci. Eng. 1 (2004), pp. 91–93.
[48] H.L. Smith and H.R. Zhu, *Stable periodic orbits for a class of three dimensional competitive systems*, J. Diff. Equations 110 (1994), pp. 143–156.
[49] R.A. Smith, *Orbital stability for ordinary differential equations*, J. Diff. Equations 69 (1987), pp. 265–287.
[50] J.P. Tian and J. Wang, *Global stability for cholera epidemic models*, Math. Biosci. 232 (2011), pp. 31–41.
[51] J.H. Tien and D.J.D. Earn, *Multiple transmission pathways and disease dynamics in a waterborne pathogen model*, Bull. Math. Biol. 72 (2010), pp. 1502–1533.
[52] V. Tudor and I. Strati, *Smallpox, Cholera*, Abacus Press, Tunbridge Wells, 1977.
[53] P. van den Driessche and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci. 180 (2002), pp. 29–48.
[54] E. Vynnycky, A. Trindall and P. Mangtani, *Estimates of the reproduction numbers of spanish influenza using morbidity data*, Int. J. Epidemiol. 36 (2007), pp. 881–889.
[55] J. Zhang and Z. Ma, *Global dynamics of an SEIR epidemic model with saturating contact rate*, Math. Biosci. 185 (2003), pp. 15–32.