Mathematical Biosciences 234 (2011) 118–126

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Global dynamics of cholera models with differential infectivity

Zhisheng Shuai *, P. van den Driessche

Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3R4

ARTICLE INFO

Article history:
Received 28 March 2011
Received in revised form 13 September 2011
Accepted 22 September 2011
Available online 2 October 2011

Keywords:
Cholera
Non-linear incidence
Differential infectivity
Global stability
Lyapunov function
Final size

ABSTRACT

A general compartmental model for cholera is formulated that incorporates two pathways of transmission, namely direct and indirect via contaminated water. Non-linear incidence, multiple stages of infection and multiple states of the pathogen are included, thus the model includes and extends cholera models in the literature. The model is analyzed by determining a basic reproduction number $\mathcal{R}_0$ and proving, by using Lyapunov functions and a graph-theoretic result based on Kirchhoff's Matrix Tree Theorem, that it determines a sharp threshold. If $\mathcal{R}_0 \leqslant 1$, then cholera dies out; whereas if $\mathcal{R}_0 > 1$, then the disease tends to a unique endemic equilibrium. When input and death are neglected, the model is used to determine a final size equation or inequality, and simulations illustrate how assumptions on cholera transmission affect the final size of an epidemic.

1. Introduction

Cholera and other waterborne diseases can be transmitted directly to humans by person-to-person contact or indirectly to humans via contaminated water. Cholera, which is a bacterial disease caused by the bacterium *Vibrio cholerae*, is a diarrheal disease that can, if untreated, lead to death. It is endemic in some regions (e.g., the Bay of Bengal [20]), but manifests as an epidemic outbreak in other regions (e.g., the Haiti outbreak in 2010 [1,41] and the outbreak starting in 2008 in Zimbabwe [31]). For more information about cholera and its occurrence, we refer the reader to [15]. Several models for this disease that take account of its two transmission pathways have been proposed and analyzed; see, for example, [1,5,7,9,14,19,31,38–41]. These models differ from each other in some aspects. Experimental studies suggest that a relatively high level of pathogen is needed to develop cholera [34], and thus saturating incidence is normally assumed for the indirect transmission [7,14,38,39], while mass action incidence is also seen in the literature (e.g., [9,40,41]). Laboratory studies also suggest that the infectivity of *Vibrio cholerae* existing outside the host decays in time, and thus cholera models with hyperinfectious and lower-infectious states of the pathogen have been studied (e.g., see [1,14]). For the direct transmission due to person-to-person contact, the differential infectivity of infectious individuals can be modeled using multiple infection stages [40]; the resulting models are usually called multi-stage models or stage progression models.

In this paper a general compartmental ordinary differential equation model for the transmission of cholera is proposed that incorporates both direct and indirect transmission, non-linear incidence, multiple infectious states of the pathogen, and multiple infection stages of infectious individuals. General forms for direct and indirect transmission include both mass action and saturating incidence. Heterogeneity in infectious host individuals and in the pathogen are included in terms of arbitrary numbers of stages of infectious individuals and states of the pathogen, respectively. Our model contains earlier cholera models in [7,14,31,38–40] as special cases, and our analysis also incorporates and extends many of the previous results. Specifically, the basic reproduction number $\mathcal{R}_0$ is determined and proved to be a sharp threshold for the model with recruitment and death: if $\mathcal{R}_0 \leqslant 1$, then the disease-free equilibrium is globally asymptotically stable, and cholera dies out; if $\mathcal{R}_0 > 1$, then the disease-free equilibrium is unstable, a unique endemic equilibrium is globally asymptotically stable, and the disease persists at a positive level. As is the case for many other complex epidemic models, proving the global stability of the endemic equilibrium imposes significant mathematical challenges. For example, global stability of the endemic equilibrium has been established for only a few cholera models using various methods such as monotone dynamical systems, the geometric approach, and the method of Lyapunov functions (see [38–40]), while such global stability results are missing for other cholera models (e.g., the hyperinfectivity model in [14] and the stage progression model in [40]). In addition, some global stability proofs (see [31, Supporting Information] and [39, Theorem 4.13]) apply a constant

* Corresponding author.
E-mail addresses: zshuai@uvic.ca (Z. Shuai), pvdd@math.uvic.ca (P. van den Driessche).

0025-5564/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2011.09.003

matrix result to a non-constant matrix; see Section 7 for further discussion. Here we successfully apply the graph-theoretic approach recently developed in [12,13,27] to construct a Lyapunov function for our cholera model and thus prove the global stability of the endemic equilibrium. Hence, for the first time, we completely establish the global dynamics for models in [14,40] and also provide a unified proof for the global stability results in [31,38–40].

The paper is organized as follows. In Section 2 we formulate our model including recruitment and death, then in Section 3 we consider equilibria and calculate the basic reproduction number $\mathcal{R}_0$. The next two sections determine global stability in the case $\mathcal{R}_0 \leqslant 1$ (Section 4) and $\mathcal{R}_0 > 1$ (Section 5). For our model without recruitment and death, in Section 6 we consider the final size of a cholera epidemic. We conclude in Section 7 with a discussion.

2. Model formulation

The total population is divided into $n+2$ compartments: a susceptible compartment, n infectious compartments representing different infection stages [40], and a removed compartment, with the number of individuals in each compartment given by S, I_i, $1 \leqslant i \leqslant n$, and R, respectively. These letters are also used to identify the compartment. A stage I_i can be interpreted as a latent stage if the infectivity of individuals in I_i is assumed to be zero. For cholera, the incubation period ranges from a few hours to 5 days, usually 2–3 days [15], although the latent compartment is normally neglected in other models cited above. The contaminated water is categorized into m levels [14], with the pathogen concentration given by B_k, $1 \leqslant k \leqslant m$. Pathogen shed from infectious individuals in each infection stage enter B_1, then progress to B_2 and so on. For cholera disease, the infectivity of B_k normally decreases as m increases [14]. Susceptible individuals can be infected either by contacting infectious individuals (direct transmission) or by ingesting contaminated water (indirect transmission). All newly infected individuals first enter the stage I_1, then I_2 and so on; see Fig. 1 for the flow diagram of this model. The incidence function is assumed to be of the form

$$\sum_{j=1}^{n} f_j(S, I_j) + \sum_{j=1}^{m} g_j(S, B_j),$$

where f_i and g_k represent direct transmission and indirect transmission, respectively. In the literature, functions f_i and g_k take different forms, such as mass action incidence $\phi(x,y) = \beta xy$ and saturating incidence $\phi(x,y) = \beta x \frac{y}{K+y}$; see Table 1. Infectious individuals I_i are assumed to contaminate the water by shedding the pathogen at a rate $h_i(I_i)$ per unit of pathogen concentration. In the literature, it is usually assumed that h_i is linear; see Table 1. Our analysis applies to the functions in Table 1 and others that satisfy the assumptions introduced below.

Based on the above assumptions and ignoring the removed individuals R (since R does not influence the dynamics of the other variables), a general cholera model can be formulated as the following system of $n+m+1$ ordinary differential equations:

$$\begin{aligned}
\frac{dS}{dt} &= A - \sum_{j=1}^{n} f_j(S, I_j) - \sum_{j=1}^{m} g_j(S, B_j) - dS,\\
\frac{dI_1}{dt} &= \sum_{j=1}^{n} f_j(S, I_j) + \sum_{j=1}^{m} g_j(S, B_j) - (d + \gamma_1 + \alpha_1) I_1,\\
\frac{dI_i}{dt} &= \gamma_{i-1} I_{i-1} - (d + \gamma_i + \alpha_i) I_i, \quad i = 2, \ldots, n,\\
\frac{dB_1}{dt} &= \sum_{j=1}^{n} h_j(I_j) - \delta_1 B_1,\\
\frac{dB_k}{dt} &= \delta_{k-1} B_{k-1} - \delta_k B_k, \quad k = 2, \ldots, m,
\end{aligned} \tag{2.1}$$

with non-negative initial conditions $S(0)$, $I_i(0)$, $B_k(0) \geqslant 0$ for all $1 \leqslant i \leqslant n$, $1 \leqslant k \leqslant m$. The removed individuals satisfy

Table 1
Special cases of model (2.1).

(n,m)	$f(S,I)$	$g(S,B)$	$h(I)$	Reference
(1,1)	0	$\frac{\lambda SB}{K+B}$	ξI	Codeço [7] Tian et al. [38]
(1,1)	0	$S\phi(B)$[a]	ξI	Tian and Wang [39]
(1,1)	βSI	$\frac{\lambda SB}{K+B}$	ξI	Tian and Wang [39] Mukandavire et al. [31]
(1,1)	βSI	λSB	$h(I)$[b]	Tian and Wang [39]
(1,1)	βSI	λSB	ξI	Tien and Earn [40]
$(n,1)$	$\beta_i SI_i$	λSB	$\xi_i I_i$	Tien and Earn [40]
(1,2)	0	$\frac{\lambda_k SB_k}{K_k+B_k}$	ξI	Hartley et al. [14]

[a] $\phi' \geqslant 0$, $\phi'' \leqslant 0$.
[b] $h' \geqslant 0$, $h'' \leqslant 0$.

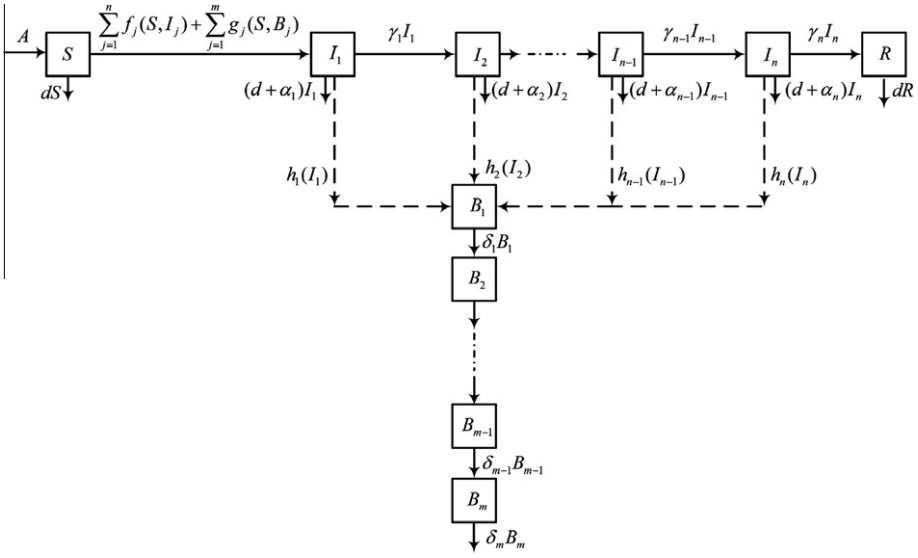

Fig. 1. The transfer diagram for model (2.1).

$$\frac{dR}{dt} = \gamma_n I_n - dR \tag{2.2}$$

with $R(0) \geqslant 0$. Here $A > 0$ represents the constant recruitment, $d > 0$ denotes the natural mortality rate, $\alpha_i \geqslant 0$, $1 \leqslant i \leqslant n$, represents the mortality rate due to the disease in the ith infection stage, $\gamma_i > 0$, $1 \leqslant i \leqslant n-1$, represents the transition rate of infectious individuals from stage I_i to I_{i+1}, $\gamma_n > 0$ represents the recovery rate of I_n, $\delta_k > 0$, $1 \leqslant k \leqslant m-1$, represents the transition rate of pathogen from B_k to B_{k+1}, and $\delta_m > 0$ represents the removal rate of B_m.

Functions f_i, g_k, and h_i are assumed to be sufficiently smooth so that solutions to (2.1) with non-negative initial conditions exist and are unique. For biological considerations, we are interested in non-negative solutions. Hence, we make the following biologically motivated assumptions throughout the paper.

- (H_1) $f_i(S, I_i) \geqslant 0$ and $f_i(S, 0) = f_i(0, I_i) = 0$ for all S, $I_i \geqslant 0$, $1 \leqslant i \leqslant n$.
- (H_2) $g_k(S, B_k) \geqslant 0$ and $g_k(S, 0) = g_k(0, B_k) = 0$ for all S, $B_k \geqslant 0$, $1 \leqslant k \leqslant m$.
- (H_3) $h_i(I_i) \geqslant 0$ and $h_i(0) = 0$ for all $I_i \geqslant 0$, $1 \leqslant i \leqslant n$.
- (H_4) There exists some index i, $1 \leqslant i \leqslant n$, such that function h_i satisfies $h_i(I_i) > 0$ for all $I_i > 0$.

Assumptions (H_1)–(H_3) ensure that solutions of (2.1) starting with non-negative initial conditions stay non-negative for all $t > 0$. Assumption (H_1) allows the possibility of $f_i \equiv 0$ for some infection stage, which can be interpreted as latent or quarantined. If all $f_i \equiv 0$, then direct transmission is ignored. Assumptions (H_3)–(H_4) allow the possibility of $h_i \equiv 0$ for some infection stage in which infectious individuals do not shed pathogen into the environment probably due to incubation or isolation, while assumption (H_4) is required to ensure involvement of the pathogen in the disease transmission.

Adding the first $n+1$ equations of (2.1) gives

$$\frac{d}{dt}(S + I_1 + \cdots + I_n) \leqslant A - d(S + I_1 + \cdots + I_n),$$

which implies that

$$\limsup_{t\to\infty}(S(t) + I_1(t) + \cdots + I_n(t)) \leqslant \frac{A}{d}.$$

Let $H_i = \max_{I_i \in [0, \frac{A}{d}]} h_i(I_i)$ and $H = \sum_{i=1}^n H_i$. It follows from the $(n+2)$th equation of (2.1) that $\frac{dB_1}{dt} \leqslant H - \delta_1 B_1$ and thus $\limsup_{t\to\infty} B_1(t) \leqslant \frac{H}{\delta_1}$. Thus from the last equation of (2.1), $\limsup_{t\to\infty} B_k(t) \leqslant \frac{H}{\delta_k}$ for all $1 \leqslant k \leqslant m$. Therefore, the feasible region

$$\Gamma = \left\{(S, I_1, \ldots, I_n, B_1, \ldots, B_m) \in \mathbb{R}_+^{n+m+1} \,|\, S + I_1 + \cdots + I_n \leqslant \frac{A}{d}, B_k \leqslant \frac{H}{\delta_k}, k = 1, \ldots, m\right\},$$

is positively invariant with respect to model (2.1).

3. Equilibria and the basic reproduction number

It follows from assumptions (H_1)–(H_3) that model (2.1) always admits a *disease-free equilibrium* (DFE) $P_0 = (S_0, 0, \ldots, 0)$ in Γ, where $S_0 = \frac{A}{d}$. Furthermore, by (H_4), P_0 is the unique equilibrium that lies on the boundary of Γ. A positive equilibrium of (2.1), if one exists, is called an *endemic equilibrium*, and denoted by $P^* = (S^*, I_1^*, \ldots, I_n^*, B_1^*, \ldots, B_m^*)$. Here $S^*, I_1^*, \ldots, I_n^*, B_1^*, B_m^* > 0$ satisfy the following equilibrium equations:

$$\begin{aligned}
A &= \sum_{j=1}^n f_j\left(S^*, I_j^*\right) + \sum_{j=1}^m g_j(S^*, B_j^*) - dS^*,\\
\mu_1 I_1^* &= \sum_{j=1}^n f_j\left(S^*, I_j^*\right) + \sum_{j=1}^m g_j\left(S^*, B_j^*\right),\\
\mu_i I_i^* &= \gamma_{i-1} I_{i-1}^*, \quad i = 2, 3, \ldots, n,\\
\delta_1 B_1^* &= \sum_{j=1}^n h_j\left(I_j^*\right),\\
\delta_k B_k^* &= \delta_{k-1} B_{k-1}^*, \quad k = 2, 3, \ldots, m,
\end{aligned} \tag{3.1}$$

where $\mu_i = d + \gamma_i + \alpha_i > 0$, $1 \leqslant i \leqslant n$.

The following assumptions are introduced to define the basic reproduction number for the system (2.1).

- (H_5) $\lim_{x\to 0^+} \frac{f_i(S_0, x)}{x} = p_i \geqslant 0$ for all $1 \leqslant i \leqslant n$.
- (H_6) $\lim_{x\to 0^+} \frac{g_k(S_0, x)}{x} = q_k \geqslant 0$ for all $1 \leqslant k \leqslant m$.
- (H_7) $\lim_{x\to 0^+} \frac{h_i(x)}{x} = r_i \geqslant 0$ for all $1 \leqslant i \leqslant n$.

Using (H_1)–(H_3), it can be easily verified that $p_i = \frac{\partial f_i}{\partial I_i}(S_0, 0)$, $q_k = \frac{\partial g_k}{\partial B_k}(S_0, 0)$, $r_i = h_i'(0)$ if f_i, g_k, h_i are differentiable.

We assume that new infections occur only in the I_1 compartment. Define two $(n+m) \times (n+m)$ matrices

$$F = \begin{pmatrix} p_1 & \cdots & p_n & q_1 & \cdots & q_m \\ 0 & \cdots & & & \cdots & 0 \\ \vdots & & & & & \\ 0 & \cdots & & & \cdots & 0 \end{pmatrix}$$

and

$$V = \begin{pmatrix} U & 0 \\ -R & D \end{pmatrix}$$

with bidiagonal

$$U = \begin{pmatrix} \mu_1 & & & & \\ -\gamma_1 & \mu_2 & & & \\ & -\gamma_2 & \ddots & & \\ & & \ddots & \mu_{n-1} & \\ & & & -\gamma_{n-1} & \mu_n \end{pmatrix},$$

$$R = \begin{pmatrix} r_1 & r_2 & \cdots & r_{n-1} & r_n \\ 0 & \cdots & & \cdots & 0 \\ \vdots & & & & \\ 0 & \cdots & & \cdots & 0 \end{pmatrix}$$

and bidiagonal

$$D = \begin{pmatrix} \delta_1 & & & & \\ -\delta_1 & \delta_2 & & & \\ & -\delta_2 & \ddots & & \\ & & \ddots & \delta_{m-1} & \\ & & & -\delta_{m-1} & \delta_m \end{pmatrix}.$$

The dimensions of U, R and D are $n \times n$, $m \times n$ and $m \times m$, respectively. Notice that both U and D have the Z-sign pattern, that is, non-positive off-diagonal entries. It can be verified that both U and D are non-singular M-matrices [3, p.137], and thus $U^{-1} \geqslant 0$, $D^{-1} \geqslant 0$. In fact, the (i,j) entries of U^{-1} and D^{-1} satisfy

$$U_{ij}^{-1} = \begin{cases} \frac{1}{\mu_i} & 1 \leqslant i = j \leqslant n, \\ 0 & 1 \leqslant i \leqslant n, 1 \leqslant i < j \leqslant n, \\ \frac{\prod_{k=j}^{i-1}\gamma_k}{\prod_{k=j}^{i}\mu_k} & 1 \leqslant i \leqslant n, \quad 1 \leqslant j < i \leqslant n \end{cases}$$

and

$$D_{ij}^{-1} = \begin{cases} \frac{1}{\delta_i} & 1 \leqslant j \leqslant i \leqslant m, \\ 0 & 1 \leqslant i < j \leqslant m, \end{cases}$$

respectively. As a consequence,

$$V^{-1} = \begin{pmatrix} U^{-1} & 0 \\ D^{-1}RU^{-1} & D^{-1} \end{pmatrix} \geqslant 0$$

and thus V is a non-singular M-matrix. Following [8,42], the basic reproduction number $\mathcal{R}_0$ is defined as the spectral radius of matrix FV^{-1}, that is,

$$\mathcal{R}_0 = \rho(FV^{-1})$$

and the DFE of (2.1) is locally asymptotically stable if $\mathcal{R}_0 < 1$, whereas it is unstable if $\mathcal{R}_0 > 1$. Since F has rank 1, straightforward calculation gives

$$\mathcal{R}_0 = \frac{p_1}{\mu_1} + \frac{p_2\gamma_1}{\mu_1\mu_2} + \cdots + \frac{p_n\gamma_1\cdots\gamma_{n-1}}{\mu_1\cdots\mu_n} + \left(\frac{q_1}{\delta_1} + \cdots + \frac{q_m}{\delta_m}\right)\left(\frac{r_1}{\mu_1} + \cdots + \frac{r_n\gamma_1\cdots\gamma_{n-1}}{\mu_1\cdots\mu_n}\right). \tag{3.2}$$

Each term involving p_i arises from direct transmission and is multiplied by the probability of reaching I_i (before dying) and the average time spent in that compartment. Each term involving q_k arises from indirect transmission and is multiplied by the average time in B_k and the sum of the probabilities of surviving to each infectious stage times the rate of pathogen shedding from that stage. This expression for $\mathcal{R}_0$ generalizes that in [14, Eq. (4)] and those in [40, Eqs. (7) and (A.7)].

4. Global dynamics when $\mathcal{R}_0 \leqslant 1$

To establish the global stability of the disease-free equilibrium, we further make several biologically reasonable assumptions on the disease transmission terms. Assume that

- (A_1) $f_i(S,I_i) \leqslant f_i(S_0,I_i) \leqslant p_iI_i$ for all $0 \leqslant S \leqslant S_0, I_i \geqslant 0$ and $1 \leqslant i \leqslant n$;
- (A_2) $g_k(S,B_k) \leqslant g_k(S_0,B_k) \leqslant q_kB_k$ for all $0 \leqslant S \leqslant S_0$, $B_k \geqslant 0$ and $1 \leqslant k \leqslant m$;
- (A_3) $h_i(I_i) \leqslant r_iI_i$ for all $I_i \geqslant 0$ and $1 \leqslant i \leqslant n$.
- (A_4) There exists either some index i, $1 \leqslant i \leqslant n$, such that function f_i satisfies $f_i(S,I_i) < f_i(S_0,I_i)$ for all $0 \leqslant S < S_0$, $I_i > 0$, or some index k, $1 \leqslant k \leqslant m$, such that function g_k satisfies $g_k(S,B_k) < g_k(S_0,B_k)$ for all $0 \leqslant S < S_0$, $B_k > 0$.

Assumptions (A_1), (A_2) and (A_4) hold for incidence functions that are monotone increasing in both variables and concave down in the second variable, including those commonly used in the literature, such as mass action and saturating incidence. Assumption (A_3) holds if function h is monotone increasing and concave down. It can be verified that all functions used in Table 1 satisfy assumptions (A_1)–(A_4).

Theorem 4.1. *Suppose that assumptions (H_1)–(H_7) hold. Then the following conclusions hold for system (2.1).*

(1) *If $\mathcal{R}_0 < 1$ and (A_1)–(A_3) hold, then the DFE is globally asymptotically stable in Γ.*
(2) *If $\mathcal{R}_0 = 1$ and (A_1)–(A_4) hold, then the DFE is globally asymptotically stable in Γ.*
(3) *If $\mathcal{R}_0 > 1$, then the DFE is unstable and system (2.1) is uniformly persistent.*

Proof. Motivated by Guo and Li [11], let $(w_1,\ldots,w_{n+m})^T = (p_1,\ldots,p_n,q_1,\ldots,q_m)^TV^{-1}$. Since $V^{-1} \geqslant 0$, it follows that $w_j \geqslant 0$ for all $1 \leqslant j \leqslant n+m$. Notice that $w_1 = \mathcal{R}_0 \leqslant 1$. Construct a Lyapunov function

$$L = w_1I_1 + \cdots + w_nI_n + w_{n+1}B_1 + \cdots + w_{n+m}B_m. \tag{4.1}$$

Differentiating L along solutions of (2.1) and using (A_1)–(A_3) yields

$$\begin{aligned} L' = \left.\frac{dL}{dt}\right|_{(2.1)} &= w_1\left(\sum_{j=1}^{n} f_j(S,I_j) + \sum_{j=1}^{m} g_j(S,B_j)\right) - \sum_{j=1}^{n} w_j\mu_jI_j \\ &\quad + \sum_{j=2}^{n} w_j\gamma_{j-1}I_{j-1} + w_{n+1}\sum_{j=1}^{n} h_j(I_j) - \sum_{j=1}^{m} w_{n+j}\delta_jB_j + \sum_{j=2}^{m} w_{n+j}\delta_{j-1}B_{j-1} \\ &\leqslant w_1\left(\sum_{j=1}^{n} p_jI_j + \sum_{j=1}^{m} q_jB_j\right) - \sum_{j=1}^{n} w_j\mu_jI_j + \sum_{j=2}^{n} w_j\gamma_{j-1}I_{j-1} \\ &\quad + w_{n+1}\sum_{j=1}^{n} r_jI_j - \sum_{j=1}^{m} w_{n+j}\delta_jB_j + \sum_{j=2}^{m} w_{n+j}\delta_{j-1}B_{j-1} \\ &= w_1(p_1,\ldots,p_n,q_1\ldots,q_m)^T(I_1,\ldots,I_n,B_1,\ldots,B_m) \\ &\quad - (w_1,\ldots,w_{n+m})^TV(I_1,\ldots,I_n,B_1,\ldots,B_m) \\ &= w_1(p_1,\ldots,p_n,q_1\ldots,q_m)^T(I_1,\ldots,I_n,B_1,\ldots,B_m) \\ &\quad - (p_1,\ldots,p_n,q_1\ldots,q_m)^T(I_1,\ldots,I_n,B_1,\ldots,B_m) \\ &= (w_1-1)\left(\sum_{j=1}^{n} p_jI_j + \sum_{j=1}^{m} q_jB_j\right) \leqslant 0, \qquad \text{if } \mathcal{R}_0 \leqslant 1. \end{aligned} \tag{4.2}$$

If $\mathcal{R}_0 < 1$, $L' = 0$ implies that $\sum_{j=1}^{n} p_jI_j + \sum_{j=1}^{m} q_jB_j = 0$, by (4.2). It follows that $\sum_{j=1}^{n} f_j(S,I_j) + \sum_{j=1}^{m} g_j(S,B_j) = 0$, which implies $S' = A - dS$ and $I_1' = -\mu_1I_1$ by the first two equations of (2.1). Hence, the invariant set where $L' = 0$ satisfies $S = S_0 = \frac{A}{d}$ and $I_1 = 0$. Similarly, the remaining equations of (2.1) give $I_2 = \cdots = I_n = 0$, $B_1 = \cdots = B_m = 0$. That is, the largest invariant set where $L' = 0$ is the singleton $\{P_0\}$. By LaSalle's Invariance Principle [25], P_0 is globally asymptotically stable in Γ if $R_0 < 1$.

If $\mathcal{R}_0 = w_1 = 1$, $L' = 0$ implies that $f_i(S,I_i) = f_i(S_0,I_i) = p_iI_i$ and $g_k(S,B_k) = g_k(S_0,B_k) = q_kB_k$ for all $1 \leqslant i \leqslant n$, $1 \leqslant k \leqslant m$. It follows from assumption (A_4) that $S = S_0$ or $I_i = B_k = 0$ for all i, k. Substituting $S = S_0$ into the first equation of (2.1) gives $\sum_{j=1}^{n} f_j(S,I_j) + \sum_{j=1}^{m} g_j(S,B_j) = 0$. Thus, as in the case $\mathcal{R}_0 < 1$, the largest invariant set where $L' = 0$ is the singleton $\{P_0\}$, and thus by LaSalle's Invariance Principle [25], P_0 is globally asymptotically stable in Γ if $R_0 = 1$.

If $\mathcal{R}_0 > 1$, then by continuity, $L' > 0$ in a neighborhood of P_0 in the interior of Γ, denoted by $\overset{\circ}{\Gamma}$. Solutions in $\overset{\circ}{\Gamma}$ sufficiently close to P_0 move away from P_0, implying that P_0 is unstable. Using a uniform persistence result from [10] and an argument as in the Proof of Proposition 3.3 of [26], it can be shown that, when $R_0 > 1$, instability of P_0 implies uniform persistence of (2.1). □

Uniform persistence of (2.1) and the positive invariance of the compact set Γ imply the existence of an equilibrium of (2.1) in $\overset{\circ}{\Gamma}$ (see Theorem D.3 in [37] or Theorem 2.8.6 in [4]).

Proposition 4.2. *Suppose that assumptions (H_1)–(H_7) hold. If $\mathcal{R}_0 > 1$, then there exists at least one endemic equilibrium for system (2.1).*

5. Global dynamics when $\mathcal{R}_0 > 1$

By Proposition 4.2, an endemic equilibrium $P^* = (S^*, I_1^*, \ldots, I_n^*, B_1^*, \ldots, B_m^*)$ exists. Here $S^*, I_1^*, \ldots, I_n^*, B_1^*, \ldots, B_m^*$ are positive and satisfy the equilibrium equations (3.1). In this section, we prove the uniqueness and global stability of P^*.

Assume that

(B_1) there exists a function $\Phi : (0, S_0] \to \mathbb{R}_+$ such that

$$(S - S^*)(\Phi(S) - \Phi(S^*)) > 0, \quad 0 < S \leqslant S_0, \quad S \neq S^*; \tag{5.1}$$

$$\left(\frac{f_i(S,I_i)\Phi(S^*)}{f_i(S^*,I_i^*)\Phi(S)} - 1\right)\left(1 - \frac{f_i(S^*,I_i^*)\Phi(S)I_i}{f_i(S,I_i)\Phi(S^*)I_i^*}\right) \leqslant 0,$$
$$0 < S \leqslant S_0, \quad I_i > 0, \quad 1 \leqslant i \leqslant n; \tag{5.2}$$

and

$$\left(\frac{g_k(S,B_k)\Phi(S^*)}{g_k(S^*,B_k^*)\Phi(S)} - 1\right)\left(1 - \frac{g_k(S^*,B_k^*)\Phi(S)B_k}{g_k(S,B_k)\Phi(S^*)B_k^*}\right) \leqslant 0,$$
$$0 < S \leqslant S_0, \quad B_k > 0, \quad 1 \leqslant k \leqslant m; \tag{5.3}$$

(B_2) for all $I_i > 0$, $1 \leqslant i \leqslant n$,

$$\left(\frac{h_i(I_i)}{h_i(I_i^*)} - 1\right)\left(1 - \frac{h_i(I_i^*)I_i}{h_i(I_i)I_i^*}\right) \leqslant 0. \tag{5.4}$$

Most functions f_i, g_k, and h_i that are commonly used in the literature, including those given in Table 1, satisfy the above assumptions (5.1)–(5.4).

Define the non-negative $(n+m)\times(n+m)$ weight matrix $W = (w_{ij})$ as

$$w_{ij} = \begin{cases} f_j(S^*, I_j^*) & i = 1, \quad 1 \leqslant j \leqslant n, \\ g_k(S^*, B_k^*) & i = 1, \quad j = n+k, \ 1 \leqslant k \leqslant m, \\ \gamma_{i-1} I_{i-1}^* & 2 \leqslant i \leqslant n, \quad j = i-1, \\ h_j(I_j^*) & i = n+1, \quad 1 \leqslant j \leqslant n, \\ \delta_{k-1} B_{k-1}^* & i = n+k, \quad 2 \leqslant k \leqslant m, \ j = i-1, \\ 0 & \text{otherwise}. \end{cases} \tag{5.5}$$

Note that if $g_m(S^*, B_m^*) > 0$ and $h_n(I_n^*) > 0$ then W is irreducible. The global stability of the endemic equilibrium is established in the following result. The proof utilizes global Lyapunov functions that are motivated by the work in [21–24] and the graph-theoretic approach for the large-scale systems recently developed in [12,13,27].

Theorem 5.1. *Suppose that assumptions (H_1)–(H_7) hold. Assume that W is irreducible and (B_1)–(B_2) hold. If $\mathcal{R}_0 > 1$, then system (2.1) has a unique endemic equilibrium P^* that is globally asymptotically stable in $\overset{\circ}{\Gamma}$.*

Proof. For system (2.1), consider the following Lyapunov function

$$V = c_1 \int_{S^*}^{S} \frac{\Phi(\xi) - \Phi(S^*)}{\Phi(\xi)} d\xi + \sum_{j=1}^{n} c_j \left(I_j - I_j^* - I_j^* \ln \frac{I_j}{I_j^*}\right) + \sum_{j=1}^{m} c_{n+j} \left(B_j - B_j^* - B_j^* \ln \frac{B_j}{B_j^*}\right). \tag{5.6}$$

Here $c_j > 0$, $j = 1, \ldots, n+m$, are constants to be specified later. Differentiating V along solutions of (2.1) and using the equilibrium equations (3.1) to simplify gives

$$\begin{aligned} V' = \left.\frac{dV}{dt}\right|_{(2.1)} &= c_1 d(S^* - S)\left(1 - \frac{\Phi(S^*)}{\Phi(S)}\right) \\ &+ c_1 \sum_{j=1}^{n} f_j(S^*, I_j^*)\left(2 - \frac{\Phi(S^*)}{\Phi(S)} + \frac{f_j(S,I_j)\Phi(S^*)}{f_j(S^*,I_j^*)\Phi(S)} - \frac{I_1}{I_1^*} - \frac{f_j(S,I_j)I_1^*}{f_j(S^*,I_j^*)I_1}\right) \\ &+ c_1 \sum_{j=1}^{m} g_j(S^*, B_j^*)\left(2 - \frac{\Phi(S^*)}{\Phi(S)} + \frac{g_j(S,B_j)\Phi(S^*)}{g_j(S^*,B_j^*)\Phi(S)} - \frac{I_1}{I_1^*} - \frac{g_j(S,B_j)I_1^*}{g_j(S^*,B_j^*)I_1}\right) \\ &+ \sum_{j=2}^{n} c_j \gamma_{j-1} I_{j-1}^* \left(\frac{I_{j-1}}{I_{j-1}^*} - \frac{I_j}{I_j^*} - \frac{I_{j-1}I_j^*}{I_{j-1}^* I_j} + 1\right) \\ &+ c_{n+1} \sum_{j=1}^{n} h_j(I_j^*)\left(\frac{h_j(I_j)}{h_j(I_j^*)} - \frac{B_1}{B_1^*} - \frac{h_j(I_j)B_1^*}{h_j(I_j^*)B_1} + 1\right) \\ &+ \sum_{j=2}^{m} c_{n+j} \delta_{j-1} B_{j-1}^* \left(\frac{B_{j-1}}{B_{j-1}^*} - \frac{B_j}{B_j^*} - \frac{B_{j-1}B_j^*}{B_{j-1}^* B_j} + 1\right). \end{aligned} \tag{5.7}$$

Notice that, by (5.1),

$$(S^* - S)\left(1 - \frac{\Phi(S^*)}{\Phi(S)}\right) \leqslant 0 \tag{5.8}$$

with equality holding if and only if $S = S^*$. Let $\Theta(x) := 1 - x + \ln x$, for $x \in (0, \infty)$. Using (5.2) and the property that $\Theta(x) \leqslant 0$ with $\Theta(x) = 0$ if and only if $x = 1$, gives

$$\begin{aligned} &2 - \frac{\Phi(S^*)}{\Phi(S)} + \frac{f_j(S,I_j)\Phi(S^*)}{f_j(S^*,I_j^*)\Phi(S)} - \frac{I_1}{I_1^*} - \frac{f_j(S,I_j)I_1^*}{f_j(S^*,I_j^*)I_1} \\ &= \left(\frac{f_j(S,I_j)\Phi(S^*)}{f_j(S^*,I_j^*)\Phi(S)} - 1\right)\left(1 - \frac{f_j(S^*,I_j^*)\Phi(S)I_j}{f_j(S,I_j)\Phi(S^*)I_j^*}\right) \\ &+ 3 - \frac{\Phi(S^*)}{\Phi(S)} - \frac{f_j(S,I_j)I_1^*}{f_j(S^*,I_j^*)I_1} - \frac{f_j(S^*,I_j^*)\Phi(S)I_j}{f_j(S,I_j)\Phi(S^*)I_j^*} - \frac{I_1}{I_1^*} + \frac{I_j}{I_j^*} \\ &\leqslant \Theta\left(\frac{\Phi(S^*)}{\Phi(S)}\right) + \Theta\left(\frac{f_j(S,I_j)I_1^*}{f_j(S^*,I_j^*)I_1}\right) + \Theta\left(\frac{f_j(S^*,I_j^*)\Phi(S)I_j}{f_j(S,I_j)\Phi(S^*)I_j^*}\right) + \frac{I_j}{I_j^*} - \ln\frac{I_j}{I_j^*} - \frac{I_1}{I_1^*} + \ln\frac{I_1}{I_1^*} \\ &\leqslant \frac{I_j}{I_j^*} - \ln\frac{I_j}{I_j^*} - \frac{I_1}{I_1^*} + \ln\frac{I_1}{I_1^*}. \end{aligned} \tag{5.9}$$

Similarly, it follows from (5.3) that

$$\begin{aligned} &2 - \frac{\Phi(S^*)}{\Phi(S)} + \frac{g_j(S,B_j)\Phi(S^*)}{g_j(S^*,B_j^*)\Phi(S)} - \frac{I_1}{I_1^*} - \frac{g_j(S,B_j)I_1^*}{g_j(S^*,B_j^*)I_1} \\ &= \left(\frac{g_j(S,B_j)\Phi(S^*)}{g_j(S^*,B_j^*)\Phi(S)} - 1\right)\left(1 - \frac{g_j(S^*,B_j^*)\Phi(S)B_j}{g_j(S,B_j)\Phi(S^*)B_j^*}\right) \\ &+ \Theta\left(\frac{\Phi(S^*)}{\Phi(S)}\right) + \Theta\left(\frac{g_j(S,B_j)I_1^*}{g_j(S^*,B_j^*)I_1}\right) + \Theta\left(\frac{g_j(S^*,B_j^*)\Phi(S)B_j}{g_j(S,B_j)\Phi(S^*)B_j^*}\right) \\ &+ \frac{B_j}{B_j^*} - \ln\frac{B_j}{B_j^*} - \frac{I_1}{I_1^*} + \ln\frac{I_1}{I_1^*} \\ &\leqslant \frac{B_j}{B_j^*} - \ln\frac{B_j}{B_j^*} - \frac{I_1}{I_1^*} + \ln\frac{I_1}{I_1^*} \end{aligned} \tag{5.10}$$

and from (5.4) that

$$\begin{aligned} \frac{h_j(I_j)}{h_j(I_j^*)} - \frac{B_1}{B_1^*} - \frac{h_j(I_j)B_1^*}{h_j(I_j^*)B_1} + 1 &= \left(\frac{h_j(I_j)}{h_j(I_j^*)} - 1\right)\left(1 - \frac{h_j(I_j^*)I_j}{h_j(I_j)I_j^*}\right) \\ &+ \Theta\left(\frac{h_j(I_j)B_1^*}{h_j(I_j^*)B_1}\right) + \Theta\left(\frac{h_j(I_j^*)I_j}{h_j(I_j)I_j^*}\right) + \frac{I_j}{I_j^*} - \ln\frac{I_j}{I_j^*} - \frac{B_1}{B_1^*} + \ln\frac{B_1}{B_1^*} \\ &\leqslant \frac{I_j}{I_j^*} - \ln\frac{I_j}{I_j^*} - \frac{B_1}{B_1^*} + \ln\frac{B_1}{B_1^*}. \end{aligned} \tag{5.11}$$

Also

$$\frac{I_{j-1}}{I^*_{j-1}}-\frac{I_j}{I^*_j}-\frac{I_{j-1}I^*_j}{I^*_{j-1}I_j}+1=\Theta\left(\frac{I_{j-1}I^*_j}{I^*_{j-1}I_j}\right)+\frac{I_{j-1}}{I^*_{j-1}}-\ln\frac{I_{j-1}}{I^*_{j-1}}-\frac{I_j}{I^*_j}+\ln\frac{I_j}{I^*_j}$$
$$\leqslant\frac{I_{j-1}}{I^*_{j-1}}-\ln\frac{I_{j-1}}{I^*_{j-1}}-\frac{I_j}{I^*_j}+\ln\frac{I_j}{I^*_j} \tag{5.12}$$

and similarly,

$$\frac{B_{j-1}}{B^*_{j-1}}-\frac{B_j}{B^*_j}-\frac{B_{j-1}B^*_j}{B^*_{j-1}B_j}+1\leqslant\frac{B_{j-1}}{B^*_{j-1}}-\ln\frac{B_{j-1}}{B^*_{j-1}}-\frac{B_j}{B^*_j}+\ln\frac{B_j}{B^*_j}. \tag{5.13}$$

Let

$$x_i=\begin{cases}I_i & 1\leqslant i\leqslant n,\\ B_k & i=n+k,\quad 1\leqslant k\leqslant m.\end{cases}$$

Combining (5.7)–(5.13) and using the notation of x_i and w_{ij} defined in (5.5) gives

$$\begin{aligned}V'&\leqslant c_1\sum_{j=1}^{n}w_{1j}\left(\frac{x_j}{x^*_j}-\ln\frac{x_j}{x^*_j}-\frac{x_1}{x^*_1}+\ln\frac{x_1}{x^*_1}\right)\\&\quad+c_1\sum_{j=n+1}^{n+m}w_{1j}\left(\frac{x_j}{x^*_j}-\ln\frac{x_j}{x^*_j}-\frac{x_1}{x^*_1}+\ln\frac{x_1}{x^*_1}\right)\\&\quad+\sum_{j=2}^{n}c_jw_{j,j-1}\left(\frac{x_{j-1}}{x^*_{j-1}}-\ln\frac{x_{j-1}}{x^*_{j-1}}-\frac{x_j}{x^*_j}+\ln\frac{x_j}{x^*_j}\right)\\&\quad+c_{n+1}\sum_{j=1}^{n}w_{n+1,j}\left(\frac{x_j}{x^*_j}-\ln\frac{x_j}{x^*_j}-\frac{x_{n+1}}{x^*_{n+1}}+\ln\frac{x_{n+1}}{x^*_{n+1}}\right)\\&\quad+\sum_{j=n+2}^{n+m}c_jw_{j,j-1}\left(\frac{x_{j-1}}{x^*_{j-1}}-\ln\frac{x_{j-1}}{x^*_{j-1}}-\frac{x_j}{x^*_j}+\ln\frac{x_j}{x^*_j}\right)\\&=\sum_{i,j=1}^{n+m}c_iw_{ij}\left(\frac{x_j}{x^*_j}-\ln\frac{x_j}{x^*_j}-\frac{x_i}{x^*_i}+\ln\frac{x_i}{x^*_i}\right).\end{aligned} \tag{5.14}$$

To show that $V'\leqslant 0$, choose constants $c_i>0$ such that (5.14) vanishes. Let $(\mathcal{G},W)$ be the weighted digraph (see Appendix A) with the weight matrix W as given in (5.5). Let $c_i=\sum_{T\in\mathbb{T}_i}w(T)$ be as given in (A.1). Since W is irreducible, it follows that $c_i>0$ for all i. Identity (A.2) from Appendix A yields

$$\sum_{i,j=1}^{n+m}c_iw_{ij}\left(\frac{x_j}{x^*_j}-\ln\frac{x_j}{x^*_j}-\frac{x_i}{x^*_i}+\ln\frac{x_i}{x^*_i}\right)\equiv 0.$$

Therefore, $V'\leqslant 0$ for all $(S,I_1,\ldots,I_n,B_1,\ldots,B_m)\in\overset{\circ}{\Gamma}$. Furthermore, $V'=0$ implies that for some constant $\lambda>0$,

$$S=S^*,\quad I_1=\lambda I^*_1,\ldots,I_n=\lambda I^*_n,\quad B_1=\lambda B^*_1,\ldots,B_m=\lambda B^*_m,$$

using properties of $\Theta(x)$ and strong connectivity of the weighted graph (G,A). Substituting these relations into the first two equations of (2.1) yields

$$0=A-dS^*-\mu_1\lambda I^*_1.$$

By the first two equations of (3.1), this last equation holds only at $\lambda=1$, namely at P^*. Therefore, the only invariant set in the set $\{V'=0\}$ is the singleton $\{P^*\}$. By LaSalle's Invariance Principle [25], P^* is globally asymptotically stable in $\overset{\circ}{\Gamma}$. As a consequence, P^* is also unique. □

In the special case that $f_i(S,I_i)=\phi(S)\psi_i(I_i)$ and $g_k(S,B_k)=\phi(S)\varphi_k(B_k)$, choose $\Phi(S)=\phi(S)$ in assumption (B_1), then conditions (5.2) and (5.3) become

$$(\psi_i(I_i)-\psi_i(I^*_i))\left(\frac{\psi_i(I_i)}{I_i}-\frac{\psi_i(I^*_i)}{I^*_i}\right)\leqslant 0,\quad I_i>0,\quad 1\leqslant i\leqslant n \tag{5.15}$$

and

$$(\varphi_k(B_k)-\varphi_k(B^*_k))\left(\frac{\varphi_k(B_k)}{B_k}-\frac{\varphi_k(B^*_k)}{B^*_k}\right)\leqslant 0,\quad B_k>0,\quad 1\leqslant k\leqslant m. \tag{5.16}$$

Therefore, assumptions (B_1)–(B_2) hold if $\phi(S)$, $\psi_i(I_i)$, $\varphi_k(B_k)$, $h_i(I_i)$ are increasing and $\frac{\phi_i(I_i)}{I_i}$, $\frac{\varphi_k(B_k)}{B_k}$, $\frac{h_i(I_i)}{I_i}$ are decreasing functions. It can be verified that all functions listed in Table 1 satisfy these assumptions, and thus our global stability result holds for those models.

The endemic equilibrium P^* can be found explicitly if $n=m=1$, with mass action incidence functions $f(S,I)=\beta SI$ and $g(S,B)=\lambda SB$, and a linear function $h(I)=\xi I$ with β, λ, ξ positive constants [40]. In this case, when $\mathcal{R}_0=\frac{\beta}{\mu}S_0+\frac{\lambda\xi}{\delta\mu}S_0>1$, system (2.1) admits an endemic equilibrium $P^*=(S^*,I^*,B^*)$, where $S^*=\frac{S_0}{\mathcal{R}_0}$, $I^*=\frac{d^*(\mathcal{R}_0-1)}{\mu}$, and $B^*=\frac{\xi I^*}{\delta}$. Theorem 5.1 shows that P^* is unique and globally asymptotically stable, as also shown in [40, Proposition 3].

6. Final size of an epidemic

In the absence of recruitment and death, model (2.1) and (2.2) can be applied to estimate the final size of a cholera epidemic. In this section, assume that $A=d=\alpha_i=0$, $f_i(S,I_i)=\beta_iSI_i$, and $h_i(I_i)=\xi_iI_i$, for all $1\leqslant i\leqslant n$. The indirect transmission $g_k(S,B_k)$ is assumed to be in the form of mass action [40] or saturating incidence [7,14,31], that is, either $g_k(S,B_k)=\lambda_kSB_k$, or $g_k(S,B_k)=\lambda_kS\frac{B_k}{K_k+B_k}$, $K_k>0$, for all $1\leqslant k\leqslant m$. Without loss of generality, assume that the total population $N=S+\sum_{i=1}^{n}I_i+R\equiv 1$. The following lemma can be proved similarly as the proof of Lemma 2 in [40]; biologically, it means that disease outbreaks eventually "burn out" in the absence of birth and death.

Lemma 6.1. *Assume that all the above assumptions hold. Then, $I_i(t)\to 0$, $B_k(t)\to 0$ for all $1\leqslant i\leqslant n$, $1\leqslant k\leqslant m$, as $t\to\infty$.*

6.1. Mass action

Assume that $g_k(S,B_k)=\lambda_kSB_k$ for all $1\leqslant k\leqslant m$. The basic reproduction number $\mathcal{R}_0$, as defined in (3.2), becomes

$$\mathcal{R}_0=\frac{\beta_1}{\gamma_1}+\frac{\beta_2}{\gamma_2}+\cdots+\frac{\beta_n}{\gamma_n}+\left(\frac{\lambda_1}{\delta_1}+\cdots+\frac{\lambda_m}{\delta_m}\right)\left(\frac{\xi_1}{\gamma_1}+\cdots+\frac{\xi_n}{\gamma_n}\right). \tag{6.1}$$

Let

$$e_i=\sum_{j=1}^{i}\frac{\beta_j}{\gamma_j}+\left(\sum_{j=1}^{i}\frac{\xi_j}{\gamma_j}\right)\cdot\left(\sum_{j=1}^{m}\frac{\lambda_j}{\delta_j}\right),\quad 1\leqslant i\leqslant n$$

and

$$\epsilon_k=\sum_{j=k}^{m}\frac{\lambda_j}{\delta_j},\quad 1\leqslant k\leqslant m.$$

Notice that $e_n=\mathcal{R}_0$. Following [29,40], set

$$Q=\ln S+\sum_{j=1}^{n-1}e_jI_{j+1}+e_nR+\sum_{j=1}^{m}\epsilon_jB_j. \tag{6.2}$$

It can be verified that the derivative of Q along the model (2.1) and (2.2) vanishes. Hence, using Lemma 6.1, it follows that

$$\ln S(\infty)+e_nR(\infty)=\ln S(0)+\sum_{j=1}^{n-1}e_jI_{j+1}(0)+e_nR(0)+\sum_{j=1}^{m}\epsilon_jB_j(0).$$

In the case that a small number of infectives or pathogen is initially introduced, that is, $S(0)\approx 1$, $I_i(0)\approx 0$, $1\leqslant i\leqslant n$, $R(0)\approx 0$, $B_k(0)\approx 0$, $1\leqslant k\leqslant m$, the final size of an epidemic satisfies the following relation

$$1 - R(\infty) = \exp(-\mathcal{R}_0 R(\infty)). \tag{6.3}$$

This relation agrees those in [29,40]. An explicit solution in terms of the Lambert function for (6.3) can be found in [29, Appendix A].

6.2. Saturating incidence

Assume that $g_k(S, B_k) = \lambda_k S \frac{B_k}{K_k + B_k}$, $K_k > 0$, for all $1 \leqslant k \leqslant m$. This indirect incidence is an attempt to model, in a continuous way, the fact that a relatively high level of pathogen is needed to develop cholera; see [19] for a discontinuous g_k used to model this fact. Then the basic reproduction number is given by (6.1) but with δ_k replaced by $\delta_k K_k$, $1 \leqslant k \leqslant m$. Defining e_i and ϵ_k as in Section 6.1 but with δ_k replaced by $\delta_k K_k$, and taking Q as in (6.2), it can be verified that the derivative of Q along the model (2.1) and (2.2) is non-negative. Hence,

$$\ln S(\infty) + e_n R(\infty) \geqslant \ln S(0) + \sum_{j=1}^{n-1} e_j I_{j+1}(0) + e_n R(0) + \sum_{j=1}^{m} \epsilon_j B_j(0).$$

In the case that a small number of infectives or pathogen is initially introduced, the final size of an epidemic satisfies the following inequality.

$$1 - R(\infty) \geqslant \exp(-\mathcal{R}_0 R(\infty)), \tag{6.4}$$

giving an upper bound on the final size $R(\infty)$.

6.3. Simulations

From Sections 6.1 and 6.2, the final size of an epidemic may differ for mass action and saturating incidence. Numerical simulation is employed to further investigate this difference created by incidence functions. Consider the case when $n = 1$ and $m = 1$. As the difference appears in the term of indirect transmission, assume $\beta = 0$, that is, ignore direct transmission. Simulations in Fig. 2 use parameters for γ, ξ, δ as in [14] (with the time unit of 1 day) and take λ and K so that the basic reproduction number $\mathcal{R}_0 = 2.7$ is the same for both incidence functions. These simulations show that mass action creates a larger outbreak in the scales of the peak prevalence $\max_t\{I(t)\}$, the peak incidence $\max_t\{\lambda S(t)B(t)\}$ or $\max_t\left\{\frac{\lambda S(t)B(t)}{K+B(t)}\right\}$, and a slightly large outbreak size $R(\infty)$, while saturating incidence leads to a slightly longer outbreak period $T = \min\{\tau : I(t) < 10^{-3}, \forall t \geqslant \tau\}$; see Table 2. Solving (6.3) with $\mathcal{R}_0 = 2.7$ yields $R(\infty) \approx 0.9156$, which approximately agrees with the outbreak size for mass action and provides an upper bound from (6.4) for the outbreak size for saturating incidence (see Table 2).

Simulations in Fig. 3 demonstrate the effect of direct/indirect transmission and heterogeneous infection stages when $\mathcal{R}_0$ is held fixed. Consider the case when $n = m = 2$ and indirect transmission is in the form of saturating incidence. Infectivity of the hyperinfectious state B_1 (freshly shed pathogen) is assumed to be 700 times the infectivity of the lower-infectious state B_2 (pathogen that have stayed in the environment for a while) [14]. Curve 1 ignores direct person-to-person transmission. Parameters for λ_i, K_i, δ_i, ξ_i, γ_i are taken from [14], giving $\mathcal{R}_0 = 15.825$. Person-to-person transmission is included in Curve 2, but equal infectivity for individuals in I_1 and I_2 is assumed ($\beta_1 = \beta_2$). Parameter values for λ_i are reduced such that direct and indirect transmission contribute equally and $\mathcal{R}_0 = 15.825$ is held fixed. Curve 3 regards I_1 as a latent stage (individuals in this stage cannot directly infect susceptible individuals but can shed pathogen into the environment) and keeps the equal contribution towards $\mathcal{R}_0 = 15.825$ from direct and indirect transmission. Simulations show that direct person-to-person transmission provides a fast route for cholera spread and produces a higher peak in prevalence $\max_t\{I_1(t) + I_2(t)\}$. Heterogeneity in infectious host individuals may decelerate the disease spread and reduce the peak in prevalence.

7. Discussion

The model (2.1) proposed in this paper is a general compartmental model for the transmission of cholera as well as other waterborne diseases. Model (2.1) incorporates both direct and indirect transmission, non-linear incidence, multiple infectious states of the pathogen, and multiple infection stages of infectious individuals. The model can serve as a general framework for modeling the spread and transmission of infectious diseases, specially diseases with multiple transmission routes. For example, both direct transmission among rabbits and indirect transmission via free-living virus particles have been observed in the spread of rabbit calicivirus disease (RCD) [2].

Model (2.1) includes as special cases many cholera models in the literature; see Table 1 for the summary. It can be easily verified that our assumptions (H_1)–(H_7), (A_1)–(A_4), and (B_1)–(B_2) hold for all models summarized in Table 1. Therefore, Theorems 4.1 and 5.1 extend the global stability results in [7,38–40] to our general model (2.1). Our results, for the first time, completely establish the global dynamics of the cholera model with a hyperinfectivity state [14], and also for the multi-stage model [40, Appendix].

The global stability results in [30, Theorem 4 in Supporting Information] and [39, Theorem 4.13] are special cases of our Theo-

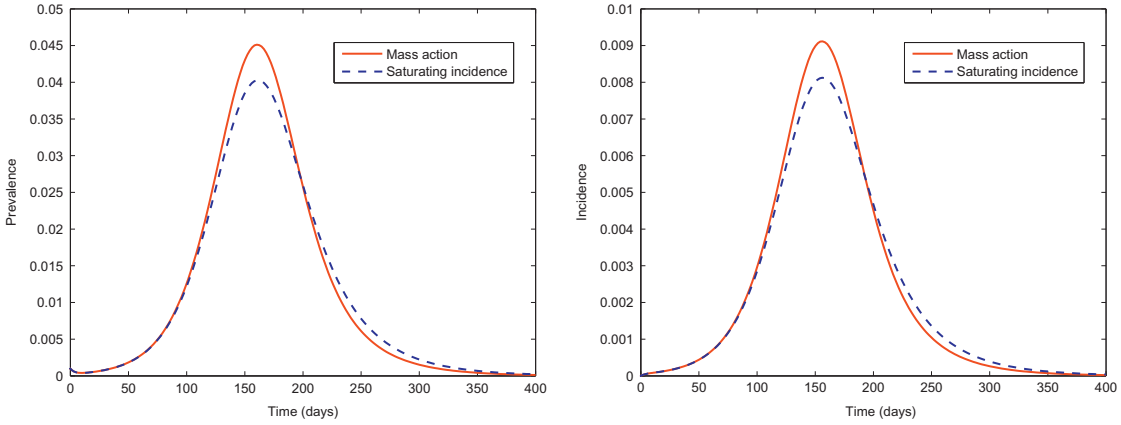

Fig. 2. The outbreak of an epidemic using models with indirect transmission given by mass action or saturating incidence. Parameter values: $\beta = 0$, $\gamma = 0.2$, $\xi = 10$, $\delta = 1/30$; for mass action $\lambda = 0.18/100$, giving $\mathcal{R}_0 = 2.7$; for saturating incidence $\lambda = 0.18$ and $K = 100$, giving $\mathcal{R}_0 = 2.7$. Initial conditions: $S(0) = 0.999$, $I(0) = 0.001$, $R(0) = B(0) = 0$ for both incidence functions.

Table 2
Summary of numerical results in Fig. 2.

	Mass action	Saturating incidence	Difference[a] (%)
Outbreak size	0.9157	0.8968	−2.1
Peak prevalence	0.0451	0.0403	−10.6
Peak incidence	0.0091	0.0081	−11.0
Outbreak period (days)	344	350	+1.7

[a] The difference is calculated by subtracting the mass action value from the saturating incidence value and dividing by the mass action value.

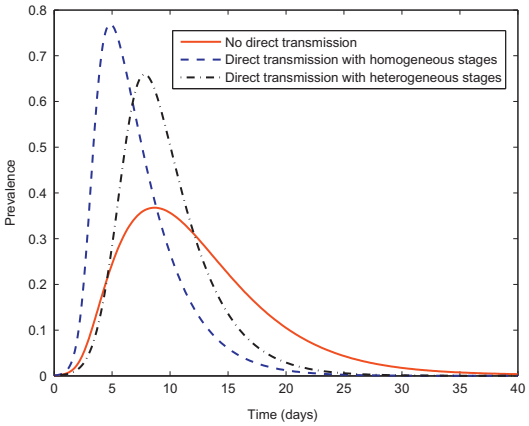

Fig. 3. The outbreak of an epidemic using model (2.1) with indirect transmission given by saturating incidence. Parameter values: $\gamma_1 = 1/2$, $\gamma_2 = 1/3$, $K_2 = 100$, $K_1 = K_2/700$, $\xi_1 = \xi_2 = 10$, $\delta_1 = 24/5$, and $\delta_2 = 1/30$. Curve 1 (no direct transmission): $\beta_1 = \beta_2 = 0$ and $\lambda_1 = \lambda_2 = 0.18$, giving $\mathcal{R}_0 = 15.825$. Curve 2 (direct transmission with homogeneous infection stages): $\beta_1 = \beta_2 = 1.5825$ and $\lambda_1 = \lambda_2 = 0.09$, giving $\mathcal{R}_0 = 15.825$. Curve 3 (direct transmission with heterogeneous infection stages): $\beta_1 = 0$, $\beta_2 = 2.6375$, and $\lambda_1 = \lambda_2 = 0.09$, giving $\mathcal{R}_0 = 15.825$. Initial conditions: $S(0) = 0.999$, $I_1(0) = 0.001$, $I_2(0) = R(0) = B_1(0) = B_2(0) = 0$ for all three cases.

rem 5.1. We note that the proofs of global stability of the endemic equilibrium in [30, Theorem 4 in Supporting Information] and [39, Theorem 4.13] are in doubt as the authors apply a constant matrix result to a nonconstant matrix. For example, the matrix A defined in [39, Eq. (4.16)] is Volterra-Lyapunov stable for each fixed pair (I,B), that is, there exists a positive 3×3 diagonal matrix $M = M(I,B) = \mathrm{diag}\{m_1(I,B),\ m_2(I,B),\ m_3(I,B)\}$ such that $MA + A^TM$ is negative definite. As a consequence, Theorem 4.12 in [39] holds for each fixed pair (I,B). However, since the positive coefficients m_i, $1 \leqslant i \leqslant 3$, possibly depend on (I,B), the derivations in Eqs. (4.12), (4.13), (4.16) in [39] may fail. Indeed the counterexample in [6] to the Markus-Yamabe conjecture has

$$A = A(x) = \begin{pmatrix} -1 & 0 & (x_1 + x_2x_3)^2 \\ -x_1 & -1 & -x_2^2x_3 - 2x_1x_2 \\ 0 & 0 & -1 \end{pmatrix},$$

and A is Volterra-Lyapunov stable for each fixed $x = (x_1,x_2,x_3)$; however, there does not exist a constant positive diagonal matrix M such that $MA + A^TM$ is negative definite. Furthermore, there exists a solution of $dx/dt = Ax$ that tends to infinity as t tends to infinity [6, Theorem 1.1].

Model (2.1) can also be regarded as a general virus dynamics model to describe the in vivo infection process of many viruses such as human immunodeficiency virus type I (HIV-I), hepatitis B virus (HBV), hepatitis C virus (HCV), and human T-cell lymphotropic virus I (HTLV-I), see [33]. In this case, the total target cells are divided into a healthy target cell compartment S, n infected target cell compartments representing different infectious stages I_i, $1 \leqslant i \leqslant n$ (see [21]), while the virions are categorized into m compartments according to the level of infectivity (see [30]). In particular, model (2.1) includes several virus dynamics models as special cases, such as [21, Eqs. (1.1) and (2.1)], [32, Eq. (1)] and [36, Eq. (1)]. Theorems 4.1 and 5.1 extend the global stability results in [21,36] to the general virus dynamics model (2.1) with non-linear incidence.

Stage progression (SP) models have been used in the literature to model the transmission and spread of infectious diseases that have a long infectious period and varying infectivity in time such as HIV/AIDS, for example, see [11,16,35]. The incidence of SP models is usually assumed to be mass action. Ignoring indirect transmission, model (2.1) becomes an SP model with general incidence functions. Global stability results Theorems 4.1 and 5.1 generalize those in [11] for mass action to general incidence functions.

One of the hypotheses in model (2.1) is that individuals recovering from cholera obtain permanent immunity, and thus model (2.1) and (2.2) is of SIR type. SIRS-type cholera models, in which removed individuals are assumed to lose immunity after a certain time and become susceptible, have also been seen in the literature, for example, see [1,9,20]. Model (2.1) and (2.2) can be easily modified to include temporary immunity and become an SIRS-type model. For the modified SIRS-type cholera model, the basic reproduction number $\mathcal{R}_0$ has the same expression as defined in (3.2); the global stability of the disease free equilibrium can be proved in the same way as the proof of Theorem 4.1 when $\mathcal{R}_0 \leqslant 1$. Biologically, the disease dies out if $\mathcal{R}_0 \leqslant 1$, and persists if $\mathcal{R}_0 > 1$. In the latter case, the uniqueness and global stability of the endemic equilibrium for the modified SIRS-type cholera model remain open.

Since the general functions in (2.1) allows the possibility of $f_i \equiv 0$ or $h_i \equiv 0$ for some infection stage i, our model can be applied to measure the effectiveness of disease control strategies such as quarantine and isolation. Vaccination can be also incorporated into (2.1) and other cholera models in [1,5,7,9,14,19,20,31,38–41] to assist public health planning to control the disease, although social and economic factors, which are not included in these models, are important for control strategies, see [17,18,28].

Acknowledgments

This research was supported in part by grants from the Natural Science and Engineering Research Council of Canada (NSERC) and from the MITACS-NCE project "Transmission Dynamics and Spatial Spread of Infectious Diseases: Modeling, Prediction and Control." Z. Shuai acknowledges the support of an NSERC Postdoctoral Fellowship. The authors thank the referees for their valuable comments and suggestions, and Drs. J.P. Tian and J. Wang for discussions.

Appendix A. Combinatorial identity

Let $(\mathcal{G}, W)$ be a weighted digraph with $N \geqslant 2$ vertices, where $W = (w_{ij})$ is the $N \times N$ weight matrix. Weight w_{ij} is positive if the directed arc (j,i) from vertex j to vertex i exists, otherwise $w_{ij} = 0$. Let $\mathbb{T}_i$ be the set of all spanning trees of $(\mathcal{G}, W)$ rooted at vertex i. For $\mathcal{T} \in \mathbb{T}_i$, the weight of $\mathcal{T}$, denoted by $w(\mathcal{T})$, is the product of weights on all arcs of $\mathcal{T}$. Let

$$c_i = \sum_{\mathcal{T} \in \mathbb{T}_i} w(\mathcal{T}), \quad i = 1, 2, \ldots, N. \tag{A.1}$$

Then $c_i \geqslant 0$, and for any family of functions $\{G_i(x_i)\}_{i=1}^N$, the following identity holds

$$\sum_{i,j=1}^{N} c_i w_{ij} G_i(x_i) = \sum_{i,j=1}^{N} c_i w_{ij} G_j(x_j). \tag{A.2}$$

If W is irreducible, then $c_i > 0$ for $i = 1,2,\ldots,N$. We refer the reader to [27] for the proof of (A.2).

References

[1] J.R. Andrews, S. Basu, Transmission dynamics and control of cholera in Haiti: an epidemic model, Lancet 377 (2011) 1248.
[2] N.D. Barlow, J.M. Kean, Simple models for the impact of rabbit calicivirus disease (RCD) on Australasian rabbits, Ecol. Model. 109 (1998) 225.
[3] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, New York, 1979.
[4] N.P. Bhatia, G.P. Szegö, Dynamical Systems: Stability Theory and Applications, Lecture Notes in Mathematics, vol. 35, Springer, Berlin, 1967.
[5] V. Capasso, S.L. Paveri-Fontana, A mathematical model for the 1973 cholera epidemic in the European Mediterranean region, Rev. Epid. San. Publ. 27 (1979) 121.
[6] A. Cima, A. van den Essen, A. Gasull, E. Hubbers, F. Mañosas, A polynomial counterexample to the Markus-Yamabe conjecture, Adv. Math. 131 (1997) 453.
[7] C.T. Codeço, Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir, BMC Infect. Dis. 1 (1) (2001).
[8] O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, J. Math. Biol. 28 (1990) 365.
[9] J.N.S. Eisenberg, M.A. Brookhart, G. Rice, M. Brown, J.M. Colford Jr., Disease transmission models for public health decision making: analysis of epidemic and endemic conditions caused by waterborne pathogens, Environ. Health Perspect. 110 (2002) 783.
[10] H.I. Freedman, M.X. Tang, S.G. Ruan, Uniform persistence and flows near a closed positively invariant set, J. Dynam. Differ. Eqn. 6 (1994) 583.
[11] H. Guo, M.Y. Li, Global dynamics of a staged progression model for infectious diseases, Math. Biosci. Eng. 3 (2006) 513.
[12] H. Guo, M.Y. Li, Z. Shuai, Global stability of the endemic equilibrium of multigroup SIR epidemic models, Can. Appl. Math. Q. 14 (2006) 259.
[13] H. Guo, M.Y. Li, Z. Shuai, A graph-theoretic approach to the method of global Lyapunov functions, Proc. Amer. Math. Soc. 136 (2008) 2793.
[14] D.M. Hartley, J.G. Morris Jr., D.L. Smith, Hyperinfectivity: a critical element in the ability of *V. cholerae* to cause epidemics?, PLOS Med 3 (2006) 63.
[15] D.L. Heymann (Ed.), Control of Communicable Diseases Manual, nineteenth ed., American Public Health Association, Washington, 2008.
[16] J.M. Hyman, J. Li, E.A. Stanley, The differential infectivity and staged progression models for the transmission of HIV, Math. Biosci. 155 (1999) 77.
[17] M. Jeuland, J. Cook, C. Poulos, J. Clemens, D. Whittington, Cost-effectiveness of new-generation oral cholera vaccines: a multisite analysis, Val. Health 12 (2009) 899.
[18] M. Jeuland, M. Lucas, J. Clemens, D. Whittington, A cost-benefit analysis of cholera vaccination programs in Beira, Mozambique, World Bank Econ. Rev. 23 (2009) 235.
[19] R.I. Joh, H. Wang, H. Weiss, J.S. Weitz, Dynamics of indirectly transmitted infectious diseases with immunological threshold, Bull. Math. Biol. 71 (2009) 845.
[20] A.A. King, E.L. Ionides, M. Pascual, M.J. Bouma, Inapparent infections and cholera dynamics, Nature 454 (2008) 877.
[21] A. Korobeinikov, Global properties of basic virus dynamics models, Bull. Math. Biol. 66 (2004) 879.
[22] A. Korobeinikov, Lyapunov functions and global stability for SIR and SIRS epidemiological models with nonlinear transmission, Bull. Math. Biol. 68 (2006) 615.
[23] A. Korobeinikov, Global properties of infectious disease models with nonlinear incidence, Bull. Math. Biol. 69 (2007) 1871.
[24] A. Korobeinikov, P.K. Maini, A Lyapunov function and global properties for SIR and SEIR epidemiological models with nonlinear incidence, Math. Biosci. Eng. 1 (2004) 57.
[25] J.P. LaSalle, The Stability of Dynamical Systems, Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1976.
[26] M.Y. Li, J.R. Graef, L. Wang, J. Karsai, Global dynamics of a SEIR model with varying total population size, Math. Biosci. 160 (1999) 191.
[27] M.Y. Li, Z. Shuai, Global-stability problems for coupled systems of differential equations on networks, J. Differ. Eqn. 248 (2010) 1.
[28] J. Lundkvist, R. Steffen, B. Jönsson, Cost-benefit of WC/rBS oral cholera vaccine for vaccination against ETEC-caused travelers' diarrhea, J. Travel Med. 16 (2009) 28.
[29] J. Ma, D.J.D. Earn, Generality of the final size formula for an epidemic of a newly invading infectious disease, Bull. Math. Biol. 68 (2006) 679.
[30] D.M. Michele, M.R. Ruy, M. Martin, D.D. Ho, A.S. Perelson, Modeling the long-term control of viremia in HIV-1 infected patients treated with antiretroviral therapy, Math. Biosci. 188 (2004) 47.
[31] Z. Mukandavire, S. Liao, J. Wang, H. Gaff, D.L. Smith, J.G. Morris Jr., Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in Zimbabwe, Proc. Natl. Acad. Sci. USA 108 (2011) 8767.
[32] M.A. Nowak, S. Bonhoeffer, A.M. Hill, R. Boehme, H.C. Thomas, H. McDade, Viral dynamics in hepatitis B virus infection, Proc. Natl. Acad. Sci. USA 93 (1996) 4398.
[33] M.A. Nowak, R.M. May, Virus Dynamics: Mathematical Principles of Immunology and Virology, Oxford University, London, 2000.
[34] T.K. Sengupta, R.K. Nandy, S. Mukhopadyay, R.H. Hall, V. Sathyamoorthy, A.C. Ghose, Characterization of a 20-k Da pilus protein expressed by a diarrheogenic strain of non-O1/non-O139 Vibrio cholerae, FEMS Microbiol. Lett. 160 (1998) 183.
[35] C.P. Simon, J.A. Jacquez, Reproduction numbers and the stability of equilibria of SI models for heterogeneous populations, SIAM J. Appl. Math. 52 (1992) 541.
[36] H.L. Smith, P. De Leenheer, Virus dynamics: a global analysis, SIAM J. Appl. Math. 63 (2003) 1313.
[37] H.L. Smith, P. Waltman, The Theory of the Chemostat: Dynamics of Microbial Competition, Cambridge University, Cambridge, 1995.
[38] J.P. Tian, S. Liao, J. Wang, Dynamical analysis and control strategies in modeling cholera, preprint. <www.math.ttu.edu/past/redraider2010/Tian2.pdf>, 2010 (accessed 14.12.10).
[39] J.P. Tian, J. Wang, Global stability for cholera epidemic models, Math. Biosci. 232 (2011) 31.
[40] J.H. Tien, D.J.D. Earn, Multiple transmission pathways and disease dynamics in a waterborne pathogen model, Bull. Math. Biol. 72 (2010) 1506.
[41] A.R. Tuite, J. Tien, M. Eisenberg, D.J.D. Earn, J. Ma, D.N. Fisman, Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions, Ann. Internal Med. 154 (2011) 593.
[42] P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, Math. Biosci. 180 (2002) 29.