



Research article

Using mathematical modeling to study the dynamics of Legionnaires' disease and consider management options

Mark Z. Wang¹, Christina J. Edholm² and Lihong Zhao^{3,*}

¹ Department of Mathematics, Pitzer College, 1050 N Mills Ave, Claremont, CA 91711, USA

² Department of Mathematics, Scripps College, 1030 N Columbia Ave, Claremont, CA 91711, USA

³ Department of Mathematics, Kennesaw State University, 1100 South Marietta Pkwy, Marietta, GA 30060, USA

* **Correspondence:** Email: lzhao12@kennesaw.edu.

Abstract: Legionnaires' disease (LD) is a largely understudied and underreported pneumonic environmentally transmitted disease caused by the bacteria *Legionella*. It primarily occurs in places with poorly maintained artificial sources of water. There is currently a lack of mathematical models on the dynamics of LD. In this paper, we formulate a novel ordinary differential equation-based susceptible-exposed-infected-recovered (SEIR) model for LD. One issue with LD is the difficulty in its detection, as the majority of countries around the world lack the proper surveillance and diagnosis methods. Thus, there is not much publicly available data or literature on LD. We use parameter estimation for our model with one of the few outbreaks with time series data from Murcia, Spain in 2001. Furthermore, we apply a global sensitivity analysis to understand the contributions of parameters to our model output. To consider managing LD outbreaks, we explore implementing sanitizing individual sources of water by constructing an optimal control problem. Using our fitted model and the optimal control problem, we analyze how different parameters and controls might help manage LD outbreaks in the future.

Keywords: Legionnaires' disease; optimal control; ordinary differential equation; mathematical model; mathematical biology; epidemiological modeling

1. Introduction

Legionnaires' disease (LD) was first identified in 1977 in the United States of America (USA) after a large pneumonia outbreak in 1976 [1]. It primarily appears in regions with poorly maintained artificial sources of water [1]. LD was responsible for 786 cases (37%) of all drinking water related illnesses in the USA from 2015 to 2020. Additionally, it accounted for the vast majority of all hospitalizations

and disease induced mortality, with 544 (97%) of all hospitalizations and 86 (98%) of all deaths [2]. In the USA, there are approximately 6000 cases reported each year, with an additional 10–15 reported cases per million population each year in Australia and Europe [1, 3]. However, due to the lack of proper surveillance and detection of LD, these numbers most likely underestimate the true number of cases. In the European Union/European Economic Area (EU/EEA) during 2021, the largest recorded rate of LD was 2.4 cases per 100,000 population [4]. While the largest recorded rate of LD by year was more recent (2021), the largest LD outbreak to date was in Murcia, Spain, in July 2001, with 449 total confirmed cases and over 800 suspected cases [5]. In this outbreak, the case-fatality rate was 1%. In this paper, we focus on a 15-day subset of this data from June 25 to July 9 [6]. During this period, there was a significant spike in the number of confirmed cases by the date of onset symptoms, which rapidly declined after July 6. In the discussed LD outbreak, the likely source of infection was from a cooling tower in the vicinity. Outbreaks such as these demonstrate the need to have proper ways to control an LD outbreak if a body of water is contaminated. This signifies the importance of sanitizing of large water sources, as the common sources of LD include artificial sources of water such as cooling towers or potable water [7].

Legionella is the bacteria responsible for LD, which is a pneumonic disease. It is transmitted through inhaling water droplets containing *Legionella pneumophila* [7]. The most frequent cause of legionellosis, *L. pneumophila*, is a specific species of *Legionella*, which causes about 90% of all reported cases of legionellosis in the USA [8]. *Legionella* is also known to cause Pontiac fever, which is a nonpneumonic disease. While its nonpneumonic counterpart Pontiac fever can go away on its own in most cases, LD requires antibiotics for treatment. The primary symptoms of Legionnaires' include fever, a loss of appetite, headache, shortness of breath, and lethargy, while health complications include shock and multi-organ failure, especially kidney and lung failure [1, 7]. Additionally, the overall disease-induced mortality rate of LD is 4%–18% through progressive pneumonia [9].

While LD surveillance has improved over time, it is still underdiagnosed and underreported. LD cannot be clinically distinguished from other pneumonic diseases [8]. Due to this difficulty, it usually requires microbiological testing for confirmation. The most common tests are generally given together: taking a culture from the lower respiratory secretion and a urinary antigen test (UAT) [10]. However, this UAT is only 70% accurate for the most common serogroup (serogroup 1) of LD [11]. It is important to improve detection since it is so difficult to separate this infection from other variations of pneumonia. This is especially necessary for high risk individuals, which include being 50 years or older, smoking or having chronic lung diseases, or having weakened immune systems [7]. From previously reported cases, about 75%–80% of cases are from individuals 50 years or older [1]. Additionally, LD is extremely costly for patients. The estimated cost per hospital stay ranges from \$7950 to \$149,000 USA dollars, let alone the productivity loss for nonfatal cases requiring hospitalization, making it the second most expensive waterborne disease in the USA to treat [9].

There have been quantitative studies performed on LD for different outbreaks. For instance, Smith et al. attempted to use data and clustering analyses to classify sources of *Legionella* which caused the cases in Genesee County, Michigan [12]. Similarly, other researchers used different data analysis techniques on LD data and incidences [13–15]. Specifically, there was an undergraduate thesis by Miramontes that formulated an initial framework of an ordinary differential equation (ODE) model for LD with high and low-risk individuals [16]. The dynamics were based on those of a disease model for

a different environmentally transmissible disease by Edholm et al. [17].

Due to the lack of mathematical models on LD, we utilized work on similar environmentally transmitted pathogens to inform our model formulation. Another relevant environmentally transmitted disease is Buruli Ulcers (BU), which is a chronic debilitating disease that causes skin damage [18]. A significant amount of past research has been done on BU that provides an approach to mathematically model an environmentally transmissible disease. Typically models for BU consider contact with the water environment that contains *Mycobacterium ulcerans*, which causes BU, or with a vector such as an insect or plant. Some of these models have used ODEs, based on the susceptible-infected-recovered (SIR) individuals framework, but incorporating the environmental transmission or other pathways [17, 19–22]. Additionally, we consider models of cholera, an environmental pathogen spread to humans by consumption or interaction with a contaminated water source [23–26]. The study of cholera gives helpful insights into understanding LD, as it is also a waterborne disease that relies on sanitization methods for control [23–25].

In this paper, building off of similar mathematical models for BU and cholera, we implement our own susceptible-exposed-infected-recovered (SEIR) individuals ODE-based model to study LD. Using the time series of new cases from the peak of the LD outbreak in Murcia, Spain, in 2001 [6], we fit our model based on the cumulative number of infections. For the parameter estimation, we do note that the focus is to capture the overall dynamics so that we can understand how model parameters influence the spread of the disease, and produce a baseline fit rather than creating a method to predict future epidemics. Therefore, based on the results from our parameter estimation, we also perform a global sensitivity analysis for the number of infected individuals. Furthermore, to provide management measures for future outbreaks of LD, we formulate an optimal control problem for our model and conduct numerical simulations of our base strategy, along with varying key parameters.

2. Model and methods

We developed a novel model for an LD outbreak that considers different compartments of individuals and a water environment containing *Legionella*. Individuals are divided into four compartments related to different stages of infection: susceptible (S), exposed (E), infected (I), and recovered (R), as reflected in Table 1. Additionally, we included a compartment for the concentration of the *Legionella* pathogen in the environment (B), since LD is contracted by exposure and interaction with the environment. Our model dynamics are captured in Figure 1. Individuals are born susceptible at the rate π and die naturally at the rate μ . Infected individuals can die from LD-related causes at the rate δ , in addition to the natural mortality rate μ . Susceptible individuals (S) may become exposed (E) by interacting with a pathogen in the environment (B) at the rate β . After spending an average of $1/\gamma$ days in the exposed state, exposed individuals (E) become infected (I). Infected individuals (I) recover at the rate λ . The system of ODEs associated with our diagram in Figure 1 is given by the following:

$$\begin{aligned}
S' &= \pi - S\beta \frac{B}{h_B + B} - \mu S, \\
E' &= S\beta \frac{B}{h_B + B} - \gamma E - \mu E, \\
I' &= \gamma E - \lambda I - \mu I - \delta I, \\
R' &= \lambda I - \mu R, \\
B' &= r_B \left(1 - \frac{B}{K_B}\right) B - d_B B.
\end{aligned} \tag{2.1}$$

The parameters are summarized in Table 2, which also includes the values we will use later for the numerical simulations. Note, in Eq (2.1), we include the logistic growth of the pathogen in the environment (B) along with natural decay to capture the concept of a carrying capacity for the concentration of the pathogen in the environment. Additionally, our choice was based on models for other environmentally transmitted pathogens, since there are no pre-existing mathematical models for *Legionella* [17, 27]. Since there is no evidence of individuals shedding *Legionella* into the environment, compartment B is not affected by the outbreak. For the interaction of susceptible individuals with the environment, we included a half-saturation function for the concentration of the pathogen in the environment (B).

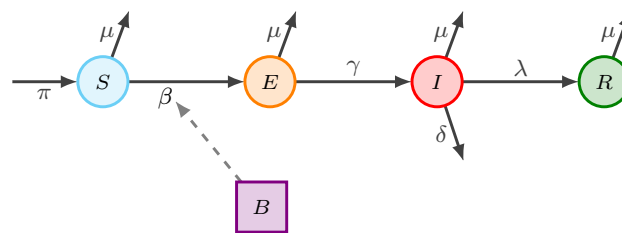


Figure 1. Flow diagram for Legionnaire's outbreak, including individuals and environment interaction. The circles represent the human compartments in the model, which flow between the classes defined by the state variables in Table 1 and parameters defined in Table 2. The square represents the concentration of the pathogen in the environment, which interacts with the transmission rate, β . The associated equations for the model are in Eq (2.1).

Table 1. Description of state variables.

| Notation | Description |
|----------|----------------------------------------------------------|
| $S(t)$ | Number of susceptible individuals at time t |
| $E(t)$ | Number of exposed individuals at time t |
| $I(t)$ | Number of infected individuals at time t |
| $R(t)$ | Number of recovered individuals at time t |
| $B(t)$ | Concentration of pathogen in the environment at time t |

Table 2. Description of parameters, along with the values used in the numerical simulations. Parameters from the literature have their associated citations, whilst parameters we estimated in Section 3.1 are indicated with an asterisk (*).

| Notation | Description | Value | Reference |
|------------|------------------------------------------------------|-----------------------|-----------|
| β | Transmission rate from S to E | 8.00×10^{-5} | * |
| h_B | Half-saturation constant of the pathogen | 199.64 | * |
| $1/\gamma$ | Average incubation period in humans | 6 | [28, 29] |
| λ | Recovery rate from infected individuals | 1/14 | [1] |
| r_B | Growth rate of the pathogen in the environment | 2.52×10^{-5} | * |
| K_B | Carrying capacity of the pathogen in the environment | 500 | * |
| d_B | Decay rate of the pathogen in the environment | 3.31×10^{-5} | * |

2.1. Basic reproduction number

For our model in Eq (2.1), we calculated the basic reproduction number, \mathcal{R}_0 , using the next generation matrix approach [30, 31]. We can establish our Jacobian matrices, F and V , from new infections \mathcal{F} and transitions \mathcal{V} for states $[E, B]^T$ at the disease-free equilibrium $(S^*, 0, 0, 0, 0) = (\frac{\pi}{\mu}, 0, 0, 0, 0)$:

$$F = \begin{pmatrix} 0 & \frac{S^*\beta}{h_B} \\ 0 & r_B \end{pmatrix}, \quad V = \begin{pmatrix} \gamma + \mu + \delta & 0 \\ 0 & d_B \end{pmatrix} \Rightarrow V^{-1} = \begin{pmatrix} \frac{1}{\gamma + \mu + \delta} & 0 \\ 0 & \frac{1}{d_B} \end{pmatrix}.$$

Therefore, to find the \mathcal{R}_0 , we take the following:

$$FV^{-1} = \begin{pmatrix} 0 & \frac{S^*\beta}{h_B} \\ 0 & r_B \end{pmatrix} \begin{pmatrix} \frac{1}{\gamma + \mu + \delta} & 0 \\ 0 & \frac{1}{d_B} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{r_B}{d_B} \end{pmatrix}.$$

Thus, $\mathcal{R}_0 = \frac{r_B}{d_B}$, which signifies that an outbreak will occur when the pathogen growth rate is larger than the natural decay rate, as described in Table 2.

2.2. Optimal control problem

We formulated an optimal control problem for our model to explore how management solutions could be applied to affect an LD outbreak. Due to the ability of *Legionella* to survive at temperature ranges between 25 and 42 °C [32], humans are presumed to be the main cause for LD outbreaks. This information suggests that the primary sources for *Legionella* growth are from artificial sources of water. Thus, we propose filtering the sources of water individuals interact with as a method of control.

For feasibility purposes, and to cut potential government expenses, we propose to clean water at the individual interaction, S with B , instead of the source, B . Thus, we introduce the time-dependent control function, $u(t)$, as the percentage ($0 \leq u(t) \leq 1$) of cleaned water, which leads to the following:

$$\begin{aligned}
S' &= \pi - (1 - u(t))S\beta\frac{B}{h_B + B} - \mu S, \\
E' &= (1 - u(t))S\beta\frac{B}{h_B + B} - \gamma E - \mu E, \\
I' &= \gamma E - (\mu + \lambda + \delta)I, \\
R' &= \lambda I - \mu R, \\
B' &= r_B\left(1 - \frac{B}{K_B}\right)B - d_B B.
\end{aligned} \tag{2.2}$$

Next, we construct the following objective functional, which represents the cost of implementing our control procedure that we want to minimize:

$$J(u, v) = \int_0^T \left[C_S(1 - u(t))S\beta\frac{B}{h_B + B} + C_u u(t) + \epsilon_u u(t)^2 \right] dt, \tag{2.3}$$

where C_S is the cost per new case, C_u represents the cost of water sanitization, ϵ_u is the nonlinear cost of water sanitization, and T denotes the final time. We set C_S to one, ϵ_u to 0.00001, and vary the sanitization cost, C_u , in our numerical simulations, with the baseline value set to 24. Since the controls, state variables, and derivatives of our state variables satisfy the boundedness, then the standard compactness results infer the existence of an optimal control function [24, 25, 33]. Using the existence result, we apply Pontryagin's Maximum Principle [34] to obtain the following optimal control characterization:

$$u(t) = \min \left\{ u_{\max}, \max \left\{ 0, \frac{\xi(\lambda_1(t) + \lambda_2(t) + C_S) - C_u}{2\epsilon_u} \right\} \right\}.$$

The Hamiltonian and adjoint equations constructed from Pontryagin's Maximum Principle are in Appendix A1.

3. Numerical simulations

3.1. Parameter estimation

We used the publicly available data of LD to approximate the recovery rate (λ) and the average incubation period ($1/\gamma$) in humans. LD treatment lasts about 1 to 3 weeks [29], while the incubation period is between 2 to 14 days, usually being 5 or 6 days [1, 28]. Hence, by averaging each time period, we set $\lambda = 1/6$ and $\gamma = 1/14$ using days as the unit of time.

For the peak of the Murcia, Spain outbreak (June 25 – July 9, 2001), there was only one death due to the disease, which proved difficult to capture with the ODE model dynamics, thus resulting in a small disease-induced mortality rate. Hence, we elected to not include this rate for the numerical simulations (i.e., $\delta = 0$). Because the simulation period was very brief (15 days), we set the natural birth and deaths rates to zero ($\pi = \mu = 0$). We used a report on the dataset related to this outbreak in which the data was given as the number of cases of LD on each date where the onset of symptoms occurred [6]. Since the data was given in this manner, fitting our parameters to the number of cumulative cases seemed more

logical, as we have data on the growing total number of cases, which is updated each day. Based on this and the amount of data, we focused on performing a general fit as a starting point for the model, understanding that we require additional information in terms of data yet to be produced for a more precise fit. To approximate the rest of the parameters, we used MATLAB's `fmincon` function from the optimization toolbox to minimize the L^2 norm of the distance between the simulated results and the data at the corresponding time points, as seen in Eq (3.1) [17, 35]:

$$F(z) = \frac{\|MI(z) - MI^*\|_2}{\|MI^*\|_2}, \quad (3.1)$$

where z is the vector of (unknown) parameters that we are estimating, $MI(z)$ is the vector with the cumulative number of infections in the numerical simulation, and MI^* is the vector with the corresponding data.

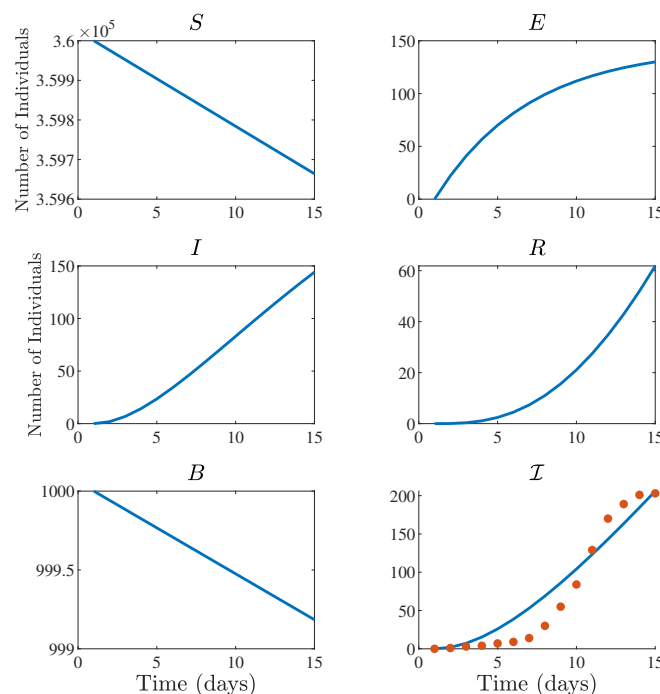


Figure 2. Time series of number of susceptible individuals (S), exposed individuals (E), infected individuals (I), recovered individuals (R), and the concentration of pathogen in the environment (B) from model, Eq (2.1), with the estimated parameter values listed in Table 2 (blue curves). The data used to estimate those parameter values are shown as orange dots in the bottom right panel for the cumulative number of infections, $I(t)$.

In our simulation, we used a total population of $N = 360,000$, and the initial conditions $E(0) = I(0) = R(0) = 0$, $S(0) = N - I(0) - R(0) - E(0) = 360,000$, $B(0) = 1000$. The resulting estimates of the parameter values are listed in the third column of Table 2, and the model simulation with these estimates is shown in Figure 2. The resulting normalized difference between the numerical simulation and the data is 0.2071. From Figure 2, we can see that the general trends of the outbreak are captured, with the individuals moving from the susceptible class, through the exposed and infected, and to recovered.

Furthermore, we can see the different dynamics of the exposed and infected classes, which are related to the incubation period and recovery rates. Our environmental class started above the carrying capacity; therefore, it is decaying towards the carrying capacity, during the short window of the outbreak related to the smaller growth and decay values associated with *Legionella*. Since we are working with a smaller dataset and a short period, our fit does a good job of capturing the general trend of the outbreak. Here, we are working to create a set of baseline parameter values for the model, which we will use to further explore their implications through a sensitivity analysis.

3.2. Sensitivity analysis

We are interested in understanding how model factors, $\theta = (\beta, h_B, \gamma, \lambda, r_B, K_B, d_B)$, affect the number of infections, I . We performed a global sensitivity analysis of the model in Eq (2.1) using Latin Hypercube Sampling (LHS) to sample the parameter space and the Partial Rank Correlation Coefficient (PRCC) to evaluate the sensitivity of the model output, $I(t)$, to variations in the model factors [36]. Each model factor is sampled 2000 times from a uniform distribution from 50% to 200% of its value listed in Table 2. As shown in Figure 3, all the parameters associated with individuals ($\beta, \gamma, \lambda, h_B$) were statistically significant, while the environmental parameters (r_B, K_B, d_B) were not; this was likely due to the assumption that the growth and decay of the pathogen in the environment were unaffected by infected individuals. Specifically, the transmission rate (β) and the exposure rate (γ) were positively correlated with the infected population I , while the recovery rate (λ) and the half-saturation constant of B (h_B) were negatively correlated with the infected population I . These correlations make sense, since increasing the transmission or exposure will result in additional infections, while decreasing the and the half-saturation means that we have less infected individuals at a given time due to a quick recovery and the diminished infection from the environment. Two parameters of specific interest to us were β and h_B , since they are related in the model dynamics for the exposure of individuals to the pathogen, as seen in Eq (2.1). The magnitude of the PRCC values for β and h_B indicate the relative importance of β in this relationship, and leads to our further exploration of this parameter with the optimal control problem in Section 3.3.

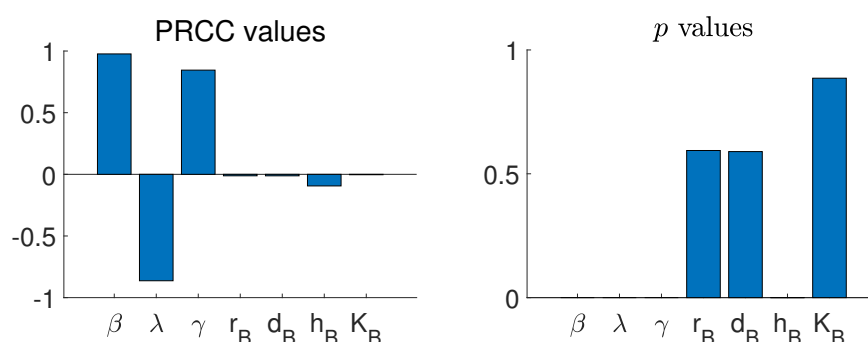


Figure 3. Global sensitivity analysis results using PRCC and LHS to examine the correlation between the number of infections, I , and the model parameters, with PRCC values in the left panel and p values in the right panel. From the p -values, we have that parameters $\beta, \gamma, \lambda, h_B$ were statistically significant, with β and γ being positively correlated with the number of infections and λ and h_B being negatively correlated.

3.3. Management method results

In Section 2.2, we introduced our objective functional in Eq (2.3) and our control vector $u(t)$ for our defined optimal control problem. The simulations were performed over $T = 20$ days to correspond with the data from our given outbreak and we set $\epsilon_u = 0.00001$. We used the forward-backward sweep method outlined in Lenhart and Workman [37] to find the optimal control vector to minimize our objective functional through numerical simulations in MATLAB. Initially, we ran the simulation with our established baseline values for the costs and parameters, with the result shown in Figure 4. We observed that the control starts at the peak allotted value for the percentage of cleaned water, assumed at 0.5; then, this transitions to no control, which makes sense with our optimal control problem formulation. Compared with Figure 2 without a control, we can see the effect of applying a control in Figure 4 especially on the exposed and infected compartments in the shape of the curves. We note that there is no change in the concentration of the pathogen in the environment B , as the proposed control method of cleaning the water from an individual interaction does not interact with the environment.

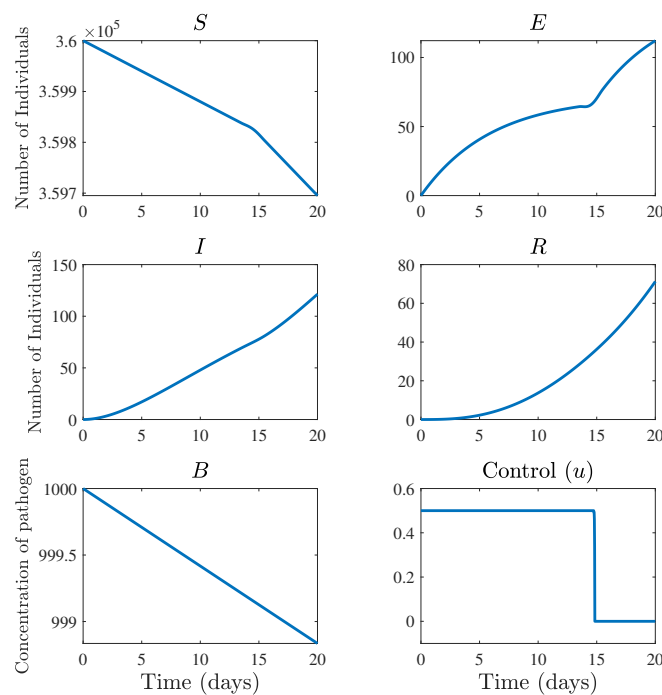


Figure 4. Numerical simulation results for the model in Eq (2.2) with the base value of $C_u = 24$, including the number of individuals that are susceptible (S), exposed (E), infected (I), and recovered (R), the concentration of pathogens in the environment (B), and the optimal level of water sanitization effort (u) over time. In comparison with Figure 2, we can observe the effect of control implementation, specifically on the exposed individuals.

We decided to explore how varying the cost of applying a sanitization control would affect the outcomes by changing the values for C_u . This would mean that the cost to apply sanitization is either lower or higher than the baseline value that was initially assumed. The results from different variations of C_u are shown in Figure 5. We varied C_u around the baseline value of 24, thereby going down to 23.98

(Variation 1) and up to 24.04 (Variation 7) by even increments. The first variation, Variation 1, results in a constant control ($u(t) \equiv 0.5$), since we have a lower cost, and will implement control throughout the 20-day outbreak. Meanwhile, our final variation, Variation 7, results in no control being applied ($u(t) \equiv 0$) since the cost of control is too high for implementation. Additional simulations confirmed that these behaviors persisted for C_u values below Variation 1 and above Variation 7. The results for Variation 2-6 show a control that is “turned on” for a portion of the outbreak, based on the cost of the control. Our base value of 24 for C_u is reflected in Variation 3, and we can see the cascade effect surrounding this as the increase and decrease in C_u result in different lengths of control implementation before dipping down to zero.

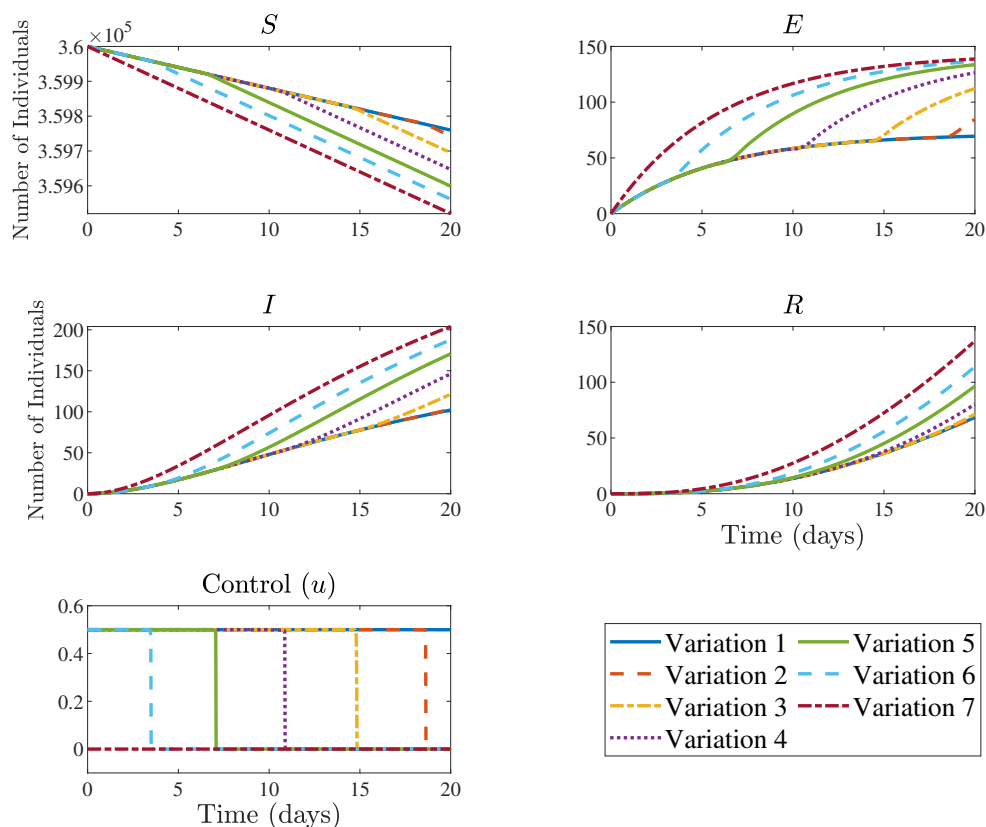


Figure 5. Numerical simulation results for the model in Eq (2.2) with different variations of C_u , the cost of water sanitization, including the number of individuals that are susceptible (S), exposed (E), infected (I), and recovered (R), and the optimal level of water sanitization effort (u) over time. The variations (1 to 7) correspond to increasing values of C_u , which demonstrate the effectiveness of the control based on cost, with a lower cost (Variation 1) leading to a longer control implementation than the high cost (Variation 7) leading to no control. Note that Variation 3 corresponds to the baseline case shown in Figure 4.

Based on our parameter estimation and sensitivity analysis, we wanted to explore the β parameter in relation to the optimal control problem. Specifically, from the sensitivity analysis, we know that the number of infected individuals is sensitive to changes in the β parameter, which we estimated and

is related to our control function. We varied the values of β up and down by 0.02% of the estimated value listed in Table 2 in even increments, and the results are shown in Figure 6. We observed a similar behavior as seen in Figure 5, though there is a different spacing to the state variables and control in these simulations. As with different variations of C_u , we observed how sensitive the control $u(t)$ is to changes in the value of β , thus highlighting the need for precise knowledge of the transmission rate. Specifically, between Variations 7 and 1 in Figure 6 for the exposed and infected individuals, the difference is around 100 individuals, which would drastically change an outbreak. Furthermore, we observed that although the evenly spaced variations of β led to evenly spaced changes in the control, $u(t)$, they did not result in the same spacing for the state variables S , E , I , and R . For instance, the difference between Variations 1 and 2 resulted in close values for the state variables for humans. As with varying C_u , when we explored values for β below Variation 1 and above Variation 7, we observed the same constant control behavior of either at the cutoff ($u(t) \equiv 0.5$) or staying at zero.

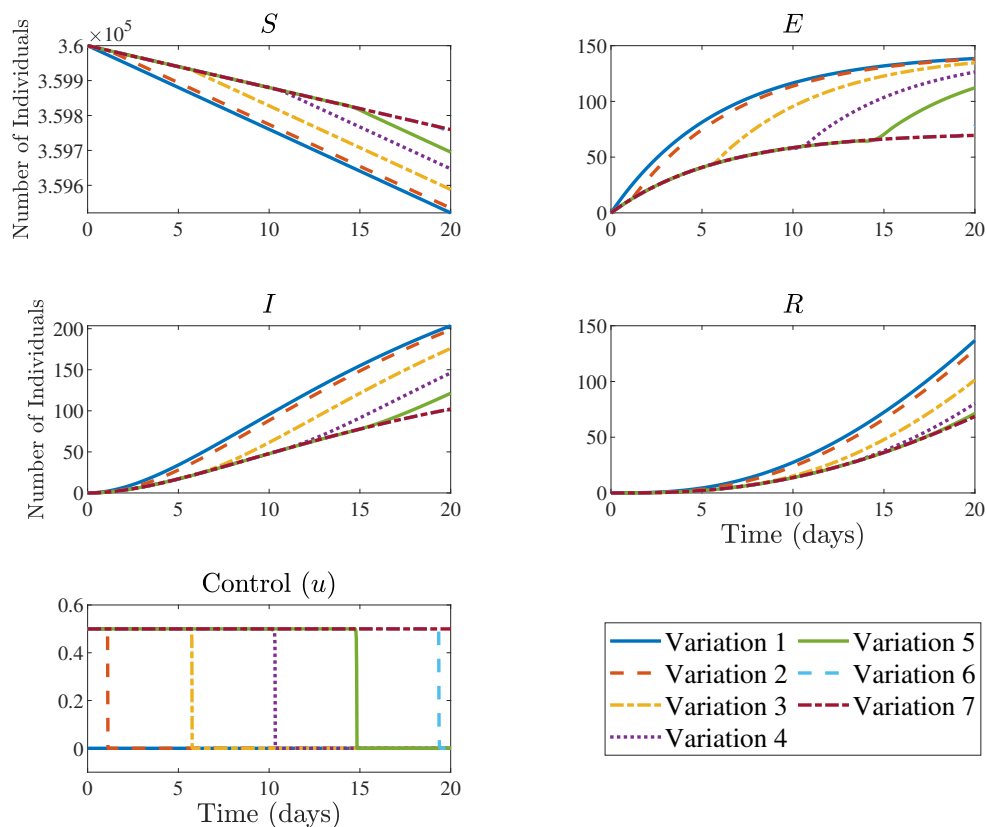


Figure 6. Numerical simulation results for the model in Eq (2.2) with different variations of β , including the number of individuals that are susceptible (S), exposed (E), infected (I), and recovered (R), and the optimal level of water sanitization effort (u) over time. The Variations (1 to 7) correspond to increasing values of β , which demonstrate the effectiveness of the control based on different values of the transmission rate. When using the lowest β in Variation 1, we apply no control due to the reduced transmission, while a high transmission in Variation 7 leads to control implementation throughout. Note that Variation 5 corresponds to the baseline case shown in Figure 4.

4. Discussion

We have formulated a novel model to capture the dynamics of LD for an outbreak. Our model formulation captures the interaction between individuals and a water environment containing the *Legionella* pathogen. For our model, we calculated the basic reproduction number and formulated an optimal control, which includes the management strategy of sanitization of individual water sources to control the spread of LD. We performed numerical simulations of our base model and optimal control problem, using parameters we estimated from a specific LD outbreak in Murcia, Spain.

From our parameter estimation, we found a baseline fit of our model to an LD outbreak that lasted for about 15 days, thus capturing the basic dynamics of an outbreak in a closed population. We decided to perform a sensitivity analysis using the LHS/PRCC global sensitivity analysis to assess the importance of various model parameters on the disease burden. From the analysis, we found that the number of infected individuals was highly sensitive to the transmission rate β , incubation period $1/\gamma$, and recovery rate λ . Meanwhile, the half-saturation rate for the environment, h_B , was statistically significant but had a very small PRCC value, which is interesting since both β and h_B are related in our model dynamics within Eq (2.1). As our optimal control problem is based on cleaning water from the individual sources, we observe that those parameters related to the pathogen in the environment (r_B , d_B , and K_B) are not all statistically significant. From these analyses, we see the value of better surveillance and data when it comes to LD outbreaks, alongside capturing the overall dynamics of the *Legionella* pathogen to incorporate the parameters related to the pathogen in the environment into our model. Based on our findings, we suggest exploring methods of decreasing the rate of transmission β , as it is a key factor that affects the change of infected cases and is one of the parameters that we can reasonably modify through sanitization. Since the primary mode that individuals contract LD is through the inhalation of this waterborne pathogen suspended in aerosolized water, the most effective way of reducing the transmission potential is through reducing the interaction with known contaminated water sources. This led to the idea of implementing sanitization of the individual sources of water in our optimal control problem.

In our optimal control problem, we used our estimated parameters to consider the implementation of sanitization management on an LD outbreak, with the baseline simulation shown in Figure 4. We noted that with sanitization, the number of exposed and infected individuals at the end of the 20-day simulation was reduced significantly. To further study the implications of management strategy accounting for the cost, we varied the cost of water sanitization, C_u , to explore its impact on the outcomes, as shown in Figure 5. Additionally, we explored the impact of varying the transmission rate β on the management strategy given the sensitivity analysis results, the results of which are summarized in Figure 6. In both scenarios, we noted the drastic changes in the number of exposed and infected individuals with slight changes in C_u or β . This was as expected, as we found that the transmission rate was highly positively correlated with the number of infected individuals. Alternatively, when we had a low cost of control, we applied control for longer lengths of time, as opposed to when the cost was increased. In general, we observed that as we had a less expensive sanitization method (i.e., low value for the cost C_u) or a more difficult outbreak to work with (i.e., high transmission rate β), we needed to implement our control longer to control the outbreak.

Due to the limitation of surveillance on LD, there is little publicly available high-quality time series data to be used to guide the model formulation. Therefore, the data used for our parameter estimation

was limited. Specifically, in the 15-day LD outbreak in Murcia, Spain [6], there are 15 data points for us to use to estimate 5 parameters in our model, which was not ideal. Additionally, since the reported data is the number of individuals on the dates where the onset of symptoms occurred, we were restricted to perform the parameter estimation against the cumulative number of infections. Thus, instead of aiming for a perfect fit, we used those estimated parameter values as a baseline poised at starting the mathematical analysis and gaining insight. Because many countries lack appropriate diagnostic, surveillance, and/or reporting systems [38], additional computational mathematical modeling and data sciences techniques could be utilized to gain insights of the true incidence of LD from the limited data available. Additionally, since there has been little to no previous effective methods of control for LD, and there is currently no known safe level of *Legionella* in water systems, we do not have a definite number for u_{\max} , which is the maximum feasible amount of sanitization percentage that we think can be cleaned in every source of water. For our model, we elected to use a higher u_{\max} in relation to cholera, though this can be further explored with additional knowledge about the feasibility of amount of water sanitized in relation to *Legionella* [24]. Additionally, there are limitations on the way we model the water environment dynamics; therefore, future work should include capturing the *Legionella* concentration in the water, B . Refining the dynamics for B will aid in better modeling the transmission, and potentially leading to preventative strategies for LD outbreaks [8].

For future directions, we are considering a variety of additional factors to calibrate our model. For instance, although *Legionella* can live in all types of water, they can thrive at 25–42 °C, with the optimal temperature for its growth being 35 °C [8]. Datasets with a monthly occurrence indicate that seasonality to be an important epidemiology feature of LD, with more cases reported during the warm season and spike in summer [32, 39]. With the trend of global warming, it would be important to explore the potential impact of climate change on LD outbreaks. Although LHS/PRCC provides information on the direction (positive or negative) of the association of model factors to the model output (number of infections in our case), it does not distinguish the contribution of single and combined factor interactions. We plan to perform a variance based global sensitivity analysis on these epidemic dynamics with the Sobol method to examine the first-order and total-order effects. While we are able to capture the dynamics of an outbreak in a closed population, our next steps include generalizing and fine-tuning our model for future outbreaks. Some potential sources of calibration include creating an age-structured model that considers individuals over the age of 50 high risk or a sex-based model that incorporates the higher risk for men. Another direction would be investigating the health inequities specific to LD, as it is known that there was a disproportionate number of infections of black individuals compared with white individuals during an LD outbreak in 2018 [40].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

We want to thank Kimi Marimontes (Scripps College 2022) and Kieran Saucedo (Harvey Mudd College 2026) for their contributions to this work in the preliminary stages.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. World Health Organization (WHO), Legionellosis, 2022. Available from: <https://www.who.int/news-room/fact-sheets/detail/legionellosis>.
2. J. M. Kunz, H. Lawinger, S. Miko, M. Gerdes, M. Thuneibat, E. Hannapel, et al., Surveillance of waterborne disease outbreaks associated with drinking water—United States, 2015–2020, *Surveillance Summ.*, **73** (2024), 1–23. <http://doi.org/10.15585/mmwr.ss7301a1>
3. Occupational Safety and Health Administration (OSHA), Legionellosis (Legionnaires' Disease and Pontiac Fever). Available from: <https://www.osha.gov/legionnaires-disease>.
4. European Centre for Disease Prevention and Control (ECDC), Legionnaires' disease - annual epidemiological report for 2021, 2023. Available from: <https://www.ecdc.europa.eu/en/publications-data/legionnaires-disease-annual-epidemiological-report-2021>.
5. A. García-Fulgueiras, C. Navarro, D. Fenoll, J. García, P. González-Diego, T. Jiménez-Buñuales, et al., Legionnaires' disease outbreak in Murcia, Spain, *Emerg. Infect. Dis.*, **9** (2003), 915–921. <http://doi.org/10.3201/eid0908.030337>
6. O. Tello, C. Pelaz, A. Garcia-Fulgueiras, C. Joseph, J. Kool, J. Lee, et al., Update on the outbreak of legionnaires' disease in Murcia, Spain, *Wkly. Releases (1997–2007)*, **5** (2001), 1714.
7. Centers for Disease Control and Prevention (CDC), Legionnaires' Disease, 2016. Available from: <https://www.cdc.gov/legionella/downloads/fs-legionnaires.pdf>.
8. B. S. Fields, R. F. Benson, R. E. Besser, Legionella and Legionnaires' disease: 25 years of investigation, *Clin. Microbiol. Rev.*, **15** (2002), 506–526. <https://doi.org/10.1128/cmr.15.3.506-526.2002>
9. D. Viasus, V. Gaia, C. Manzur-Barbur, J. Carratalà, Legionnaires' disease: Update on diagnosis and treatment, *Infect. Dis. Ther.*, **11** (2022), 973–986. <https://doi.org/10.1007/s40121-022-00635-7>
10. Centers for Disease Control and Prevention (CDC), Laboratory testing for *Legionella*, 2024. Available from: <https://www.cdc.gov/legionella/php/laboratories/index.html>.
11. J. W. Den Boer, E. P. F. Yzerman, Diagnosis of Legionella infection in Legionnaires' disease, *Eur. J. Clin. Microbiol. Infect. Dis.*, **23** (2004), 871–878. <https://doi.org/10.1007/s10096-004-1248-8>
12. A. F. Smith, A. Huss, S. Dorevitch, L. Heijnen, V. H. Arntzen, M. Davies, et al., Multiple sources of the outbreak of Legionnaires' disease in Genesee County, Michigan, in 2014 and 2015, *Environ. Health Perspect.*, **127** (2019), 127001. <https://doi.org/10.1289/EHP5663>
13. K. A. Hamilton, C. N. Haas, Critical review of mathematical approaches for quantitative microbial risk assessment (QMRA) of legionella in engineered water systems: Research gaps and a new framework, *Environ. Sci. Water Res. Technol.*, **2** (2016), 599–613. <https://doi.org/10.1039/C6EW00023A>

14. B. Prasad, K. A. Hamilton, C. N. Haas, Incorporating time-dose-response into legionella outbreak models, *Risk Anal.*, **37** (2017), 291–304. <https://doi.org/10.1111/risa.12630>
15. S. Zahran, S. P. McElmurry, P. E. Kilgore, D. Mushinski, J. Press, N. G. Love, et al., Assessment of the Legionnaires' disease outbreak in Flint, Michigan, *PNAS*, **115** (2018), E1730–E1739. <https://doi.org/10.1073/pnas.1718679115>
16. K. Miramontes, *An Examination of Legionnaires' Disease Through the Lens of ODE Modeling*, Scripps Senior Theses, 2022. Available from: https://scholarship.claremont.edu/scripps_theses/1897.
17. C. Edholm, B. Levy, A. Abebe, T. Marijani, S. Le Fevre, S. Lenhart, et al., A risk-structured mathematical model of buruli ulcer disease in Ghana, in *Mathematics of Planet Earth*, Springer, **5** (2019), 109–128. https://doi.org/10.1007/978-3-030-22044-0_5
18. World Health Organization (WHO), Buruli ulcer (Mycobacterium ulcerans infection), 2023. Available from: [https://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-\(mycobacterium-ulcerans-infection\)](https://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-(mycobacterium-ulcerans-infection)).
19. E. Bonyah, I. Dontwi, F. Nyabadza, A theoretical model for the transmission dynamics of the Buruli ulcer with saturated treatment, *Comput. Math. Methods Med.*, **2014** (2014), 576039. <https://doi.org/10.1155/2014/576039>
20. A. Garchitorea, C. N. Ngonghala, G. Texier, J. Landier, S. Eyangoh, M. H. Bonds, et al., Environmental transmission of Mycobacterium ulcerans drives dynamics of Buruli ulcer in endemic regions of Cameroon, *Sci. Rep.*, **5** (2015), 18055. <https://doi.org/10.1038/srep18055>
21. F. Nyabadza, E. Bonyah, On the transmission dynamics of Buruli ulcer in Ghana: Insights through a mathematical model, *BMC Res. Notes*, **8** (2015), 656. <https://doi.org/10.1186/s13104-015-1619-5>
22. C. Nyarko, P. Nyarko, I. Ampofi, E. Asante, Modelling transmission of Buruli Ulcer in the central region of Ghana, *Math. Modell. Appl.*, **5** (2020), 221–230.
23. E. Che, E. Numfor, S. Lenhart, A. A. Yakubu, Mathematical modeling of the influence of cultural practices on cholera infections in Cameroon, *Math. Biosci. Eng.*, **18** (2021), 8374–8391. <https://doi.org/10.3934/mbe.2021415>
24. E. Howerton, K. Dahlin, C. J. Edholm, L. Fox, M. Reynolds, B. Hollingsworth, et al., The effect of governance structures on optimal control of two-patch epidemic models, *J. Math. Biol.*, **87** (2023), 74. <https://doi.org/10.1007/s00285-023-02001-8>
25. M. R. Kelly Jr., J. H. Tien, M. C. Eisenberg, S. Lenhart, The impact of spatial arrangements on epidemic disease dynamics and intervention strategies, *J. Biol. Dyn.*, **10** (2016), 222–249. <https://doi.org/10.1080/17513758.2016.1156172>
26. J. H. Tien, D. J. D. Earn, Multiple transmission pathways and disease dynamics in a waterborne pathogen model, *Bull. Math. Biol.*, **72** (2010), 1506–1533. <https://doi.org/10.1007/s11538-010-9507-6>
27. X. Wang, J. Wang, Analysis of cholera epidemics with bacterial growth and spatial movement, *J. Biol. Dyn.*, **9** (2015), 233–261. <https://doi.org/10.1080/17513758.2014.974696>

28. J. R. Egan, I. M. Hall, D. J. Lemon, S. Leach, Modeling Legionnaires' disease outbreaks: Estimating the timing of an aerosolized release using symptom-onset dates, *Epidemiology*, **22** (2011), 188–198. <https://doi.org/10.1097/EDE.0b013e31820937c6>
29. The National Health Service (NHS), Legionnaires' disease, 2023. Available from: <https://www.nhs.uk/conditions/legionnaires-disease/>.
30. P. Van den Driessche, J. Watmough, Further notes on the basic reproduction number, in *Mathematical Epidemiology*, Springer, **1945** (2008), 159–178. https://doi.org/10.1007/978-3-540-78911-6_6
31. P. Van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Math. Biosci.*, **180** (2002), 29–48. [https://doi.org/10.1016/S0025-5564\(02\)00108-6](https://doi.org/10.1016/S0025-5564(02)00108-6)
32. T. Lupia, S. Corcione, N. Shbaklo, B. Rizzello, I. De Benedetto, E. Concialdi, et al., *Legionella pneumophila* infections during a 7-year retrospective analysis (2016–2022): Epidemiological, clinical features and outcomes in patients with Legionnaires' disease, *Microorganisms*, **11** (2023), 498. <https://doi.org/10.3390/microorganisms11020498>
33. W. H. Fleming, R. Rishel, *Optimal Deterministic and Stochastic Control*, Springer, 1975.
34. L. S. Pontryagin, V. G. Boltyanskiy, R. V. Gamkrelidze, E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Wiley, 1962.
35. C. J. Edholm, B. Levy, L. Spence, F. B. Augusto, F. Chirove, C. W. Chukwu, et al., A vaccination model for COVID-19 in Gauteng, South Africa, *Infect. Dis. Modell.*, **7** (2022), 333–345. <https://doi.org/10.1016/j.idm.2022.06.002>
36. S. Marino, I. B. Hogue, C. J. Ray, D. E. Kirschner, A methodology for performing global uncertainty and sensitivity analysis in systems biology, *J. Theor. Biol.*, **254** (2008), 178–196. <https://doi.org/10.1016/j.jtbi.2008.04.011>
37. S. Lenhart, J. T. Workman, *Optimal Control Applied to Biological Models*, Chapman and Hall/CRC, 2007. <https://doi.org/10.1201/9781420011418>
38. N. Phin, F. Parry-Ford, T. Harrison, H. R. Stagg, N. Zhang, K. Kumar, et al., Epidemiology and clinical management of Legionnaires' disease, *Lancet Infect. Dis.*, **14** (2014), 1011–1021. [https://doi.org/10.1016/S1473-3099\(14\)70713-3](https://doi.org/10.1016/S1473-3099(14)70713-3)
39. M. Riccò, S. Peruzzi, S. Ranzieri, P. G. Giuri, Epidemiology of Legionnaires' disease in Italy, 2004–2019: A summary of available evidence, *Microorganisms*, **9** (2021), 2180. <https://doi.org/10.3390/microorganisms9112180>
40. C. M. Hunter, S. W. Salandy, J. C. Smith, C. Edens, B. Hubbard, Racial disparities in incidence of Legionnaires' disease and social determinants of health: A narrative review, *Public Health Rep.*, **137** (2022), 660–671. <https://doi.org/10.1177/00333549211026781>

Appendix

A1. Details on optimal control formulation

We construct the Hamiltonian as follows:

$$\begin{aligned} H = & C_S(1-u)S\beta\frac{B}{h_B+B} + C_u u + \epsilon_u u^2 + \lambda_1 \left(\pi - (1-u)S\beta\frac{B}{h_B+B} - \mu S \right) \\ & + \lambda_2 \left((1-u)S\beta\frac{B}{h_B+B} - \gamma E - \mu E \right) \\ & + \lambda_3 (\gamma E - (\mu + \lambda + \delta)I) \\ & + \lambda_4 (\lambda I - \mu R) \\ & + \lambda_5 \left(r_B \left(1 - \frac{B}{K_B} \right) B - d_B B \right), \end{aligned}$$

and obtain our adjoint equations as follows:

$$\begin{aligned} \lambda'_1 &= \frac{\partial H}{\partial S} = - \left[C_S(1-u)\beta\frac{B}{h_B+B} - \lambda_1(1-u)\beta\frac{B}{h_B+B} - \lambda_1\mu - \lambda_2(1-u)\beta\frac{B}{h_B+B} \right], \\ \lambda'_2 &= \frac{\partial H}{\partial E} = - [\lambda_2(-\gamma - \mu) + \lambda_3\gamma], \\ \lambda'_3 &= \frac{\partial H}{\partial I} = - [\lambda_3(-\mu - \lambda - \delta) + \lambda_4\lambda], \\ \lambda'_4 &= \frac{\partial H}{\partial R} = - [-\lambda_4\mu], \\ \lambda'_5 &= \frac{\partial H}{\partial B} = - \left[C_S(1-u)S\beta\frac{h_B}{(h_B+B)^2} - \lambda_1(1-u)S\beta\frac{h_B}{(h_B+B)^2} \right. \\ & \quad \left. + \lambda_2(1-u)S\beta\frac{h_B}{(h_B+B)^2} + \lambda_5 \left(r_B \left(1 - \frac{2B}{K_B} \right) - d_B \right) \right]. \end{aligned}$$

Finally, we solve the optimality condition as follows:

$$\begin{aligned} \frac{\partial H}{\partial u} &= 0, \\ 0 &= -C_S\xi + C_u + 2\epsilon_u u - \lambda_1\xi - \lambda_2\xi, \\ u &= \frac{\xi(\lambda_1 + \lambda_2 + C_S) - C_u}{2\epsilon_u}, \end{aligned}$$

where $\xi = S\beta\frac{B}{h_B+B}$.



AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)