OPTIMAL CONTROL APPLIED TO VACCINATION AND TREATMENT STRATEGIES FOR VARIOUS EPIDEMIOLOGICAL MODELS

Holly Gaff

Community and Environmental Health, College of Health Sciences
Old Dominion University, 3133A Technology Building, Norfolk, VA 23529, USA

Elsa Schaefer

Department of Mathematics
Marymount University, 2807 North Glebe Road, Arlington, VA 22207, USA

(Communicated by Urszula Ledzewicz)

Abstract. Mathematical models provide a powerful tool for investigating the dynamics and control of infectious diseases, but quantifying the underlying epidemic structure can be challenging especially for new and under-studied diseases. Variations of standard SIR, SIRS, and SEIR epidemiological models are considered to determine the sensitivity of these models to various parameter values that may not be fully known when the models are used to investigate emerging diseases. Optimal control theory is applied to suggest the most effective mitigation strategy to minimize the number of individuals who become infected in the course of an infection while efficiently balancing vaccination and treatment applied to the models with various cost scenarios. The optimal control simulations suggest that regardless of the particular epidemiological structure and of the comparative cost of mitigation strategies, vaccination, if available, would be a crucial piece of any intervention plan.

1. **Introduction.** Accidental and intentional introduction of infectious diseases to previously naïve geographic regions has brought more focus and attention to the development of response plans to such scenarios. All levels of government and public health officials are searching for answers to identify the best strategies for intervention prior to what may be an inevitable event. Given our global connectedness, diseases that were previously isolated to various parts of the world have spread even across oceans [4].

In developing response plans to disease outbreaks, decision makers are often faced with trade-offs in choosing among various treatment options, quarantine, and/or vaccines. For example, cholera, human papilloma virus, and influenza all have approved medical treatment options as well as available vaccines. The challenge is to find the optimal response balancing treatment and vaccination that will minimize incidence and disease-related mortality while being mindful of the costs of each mitigation strategy.

Mathematical models have been used to explore the dynamics of diseases since Bernoulli used a simple model to estimate smallpox mortality and argue for the

advantages of inoculation [3]. Most mathematical models for infectious diseases use a similar underlying methodology based on the work formalized by Anderson and May [1]. The first step is to represent the epidemiology of the disease being studied by dividing the population into subpopulations, called compartments, that represent the various stages of disease progression. For example, individuals are identified as "susceptible" (S) to a disease if they don't currently have the disease nor any immunity to the disease, e.g., they have not been vaccinated. Individuals are "exposed" (E) if they have been infected with the disease pathogen but are not able to infect others, and they are "infectious" (I) if they are infected and infectious with the disease pathogen. Finally, they are "removed" (R) if they have cleared the infection and have immunity to recurrence for at least some period of time, have been successfully vaccinated against the disease, or are otherwise isolated from the population so that they cannot spread the disease. Using these compartments and possibly additional ones as needed, the epidemiology of a disease is represented as a series of subpopulations connected by the flow from one compartment to another that is dictated by the disease dynamics. The rates of flow between compartments are estimated from experiments and field data.

Hethcote provides several notable surveys of basic epidemiological models [13, 16], and he and other authors have considered in detail the effects of variations on these models. In addition, there have been several articles considering optimal control applied to specific diseases, for example [18, 19, 22, 31]. Hethcote suggests that the most needed contributions to the literature would contrast the effects of altering the basic components of the various models [13].

Brauer considers several less traditional variations, including compartments for asymptomatic, quarantined, and isolated individuals [5]. In considering the effect of treatment in these models, he finds that asymptotically the models' qualitative behaviors agree with those of more simple models. However, he notes that it would be of interest to consider the trade-off between vaccination and treatment and how the more complicated structures might influence those results.

Behncke applies optimal control to a number of general epidemiological models with general terms allowing many variations of those models [2]. The article deals mostly with the costs and timely application of vaccination to various epidemiological structures, and provides clear theoretical results in those cases.

Several key components of such models are those that consider mitigation strategies such as treatment and vaccination. Optimal control theory provides a valuable tool to begin to assess the trade-offs between vaccination and treatment strategies. Optimal control is a mathematical technique derived from the calculus of variations. There are a number of different methods for calculating the optimal control for a specific mathematical model. For example, Pontryagin's maximum principle allows the calculation of the optimal control for an ordinary differential equation model system with a given constraint. Variations of Pontryagin's maximum principle have been derived for other types of models including partial differential equations and difference equations [20, 23]. These techniques are powerful when applied to disease models and can provide great insights into the best pathway to reduce disease burden. For example, with a given mathematical model for a disease, one can calculate the best vaccination schedule balancing the cost of the vaccine and the cost of the disease burden [15].

In using optimal control to suggest an intervention strategy for a specific disease, one of the first steps is to identify the epidemic structure of a disease. During an

emerging infectious disease, it is often unclear what the exact epidemic structure should be, as can be seen from the variety of models created for the 2002-2003 SARS epidemic [24, 29, 33]. Even for existing diseases, there is often a lack of agreement on whether a given disease has, for example, a significant incubation period prior to the individual being infectious as is the case for smallpox [8, 21]. These types of questions are vital to the understanding of the disease dynamics in general. The process of developing a mathematical model forces quantitative thinking that often highlights gaps in data and current knowledge that need to be addressed.

The question we study in this paper is whether this underlying epidemic structure significantly impacts the predicted optimal control strategy for administering vaccination and/or treatment. This is an important question given that often at the beginning of a newly emerged or previously under-studied disease, the underlying epidemic structure is unknown. If the optimal control strategy remains qualitatively the same regardless of the underlying epidemic structure, then less time needs to be spent investigating that structure before initial response policies can be made. Additional research will be needed subsequently to understand the disease more fully and assess and update the response policy as needed. We present results from the exploration of SIR, SIRS, and SEIR models including vaccination and treatment options under varying rates for incidence and disease related death. Below we present the SIR model analysis in detail including the proof of uniqueness and existence of the optimal control solutions followed by summaries for the results of the similar analyses of the SIRS and SEIR models. Finally, we present the results of numerical simulations for each model under various parameter values.

2. **Equations for a standard SIR model.** We begin our explorations with a standard SIR model for a population with underlying logistic growth as defined by the following state equations:

$$\frac{dS}{dt} = \mu N - \beta \frac{SI}{N} - \nu S - \mu \frac{NS}{K}, \tag{1}$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - (\gamma + \tau + \delta)I - \mu \frac{NI}{K}, \tag{2}$$

$$\frac{dR}{dt} = (\gamma + \tau)I + \nu S - \mu \frac{NR}{K}, \tag{3}$$

subject to the boundary conditions

$$S(0) = S_0, I(0) = I_0, R(0) = R_0. \tag{4}$$

As described in the previous section, the variables S, I, and R represent the susceptible, infectious, and removed classes, respectively. We assume logistic growth of the total population $N = S + I + R$ with carrying capacity K, and we additionally assume that all new births enter the susceptible class. Note that many models treat the population size as constant, and in the short term the models provide essentially the same results. However, when considering disease-related deaths [13] the logistic model should provide more accurate predictions. It should be noted that the direct applicability of a logistic growth model to a human population remains a subject of debate [27], as human birth and death rates are far more complicated than simple resource allocation. Nonetheless, the logistic model, which provides mathematical stability, will produce reasonable results for our generic model population over a time frame that is significantly shorter than a human lifespan as is used in this optimal control study.

The parameter β approximates the average number of contacts with infectious individuals needed to make one person ill in each unit of time. Note that $\frac{I}{N}$ approximates the fraction of the population that is infectious. If we assume that each person has one contact in each time period (measured in days for our analysis), and thus the term $\frac{SI}{N}$ approximates the number of susceptible/infectious contacts in each time period, and $\beta\frac{SI}{N}$, the "standard incidence," approximates the number of susceptibles who become infectious in each time period [13]. Liu, Hethcote, and others have published multiple studies examining the need for and effect of other measures of incidence [12, 14, 17, 25, 26]. In particular, if I is raised to a power, interesting dynamics can be observed. As these researchers write, it is difficult to assess in a particular case which incidence model is most appropriate.

We assume that some proportion δ of the infectious population dies due to the disease in each time period. The control function $\nu(t)$ measures the rate at which susceptibles are vaccinated in each time period, and the control function $\tau(t)$ measures the rate at which infectious individuals are treated in each time period. Note that in equations (1) – (3) the control ν moves individuals from the S class to the R class, and the control τ moves individuals from the I class to the R class. Since quarantine effectively removes infectious individuals from the population, the control τ can provide a rudimentary model for quarantine effects in addition to modeling medical treatments of the disease. This observation may be useful for determining mitigation strategies for diseases for which no treatment is known. Note that an alternative model for quarantine introduced by Hethcote et al. adjusts the incidence term to account for a smaller mixing population, which can produce periodic disease dynamics for some parameter sets [12]. In addition, as mentioned earlier, Brauer's recent model also distinguishes between quarantine and isolation, and also adjusts other key parameters for these populations [5].

A summary of parameter meanings and values used in our numerical computations appears in Table 1. Note that for these simulations the total initial population is at carrying capacity, and the models' logistic population dynamics allow the the population to recover its size as disease-related deaths are permitted. Also, note that initial conditions have been chosen to allow our optimal control pairs to be well-defined in equation (14) that will be derived in Section 4 following.

Given initial population sizes S_0, I_0, and R_0, we seek the best mitigation strategy for the outbreak modeled in equations (1) – (3) by optimally defining bounded, *Lebesque* integrable control functions $\nu(t)$ and $\tau(t)$. Our goal is to minimize the number of people who become infected, and thus also the number of people who die due to the infection, while also minimizing the effort of vaccinating and treating the population. Thus, we seek to minimize the objective functional

$$J(\nu, \tau) = \in_0^T \left[B_1 I(t) + B_2 \left[\frac{R(t)}{K} \right]^m \nu^2(t) + B_3 \tau^2(t) \right] dt, \qquad (5)$$

where $m \geq 1$. The constants B_1, B_2, and B_3 have a dual role. On one hand, they are needed to balance the units in the integrand because the number of infected individuals will be measured in the hundreds in this paper, while τ and ν are treatment rates and will necessarily lie between 0 and 1. On the other hand, in numerical runs we further vary the constants to place stronger importance on the minimization of infected individuals or on treatment and/or vaccination efforts. The terms $B_1 I(t)$ and $B_3 \tau^2(t)$ are rather standard terms in objective functionals with similar goals (see, for example, [18, 19], or, for a contrasting objective functional

TABLE 1. Summary of parameter values used in figures

Name	Description	Value
S_0	Initial susceptible population	4500 humans
E_0	Initial exposed population	498 humans
I_0	Initial infected population	499 humans (SEIR: 1 human)
R_0	Initial removed population	1 human
K	Carrying Capacity	5000
μ	intrinsic growth rate	0.00004 / day
δ	death rate due to disease	0 to 0.1 / day
β	incidence rate	0.05 to 0.55 / day
γ	infect time	0.1 / day
ω	waning rate	0.001 / day
ϵ	transition rate	0.1 / day
B_1	weight for number infected	1
B_2	weight for vaccination	1000 to 100,000
B_3	weight for treatment	1000 to 10,000
$\nu_{\max}$	Max Vacc Rate	0.1 / day
$\tau_{\max}$	Max Treatment Rate	0.6 / day

[22, 30]). Treatment is viewed as a nonlinear function since implementation of any public health intervention does not have a linear cost, but rather, there are increasing costs with reaching higher fractions of the population. The quadratic is the most simple nonlinear function and so is used for treatment.

The remaining term in our objective functional seeks to increase the expense of vaccination when most of the population has either been vaccinated or has immunity from a prior infection. When m is chosen to be an even integer such as 10, the term $B_2[\frac{R}{K}]^m \nu^2(t)$ reflects the reality that the cost of vaccinating the first approximately 80 percent of a population is relatively small as compared with the cost of vaccinating the remaining 20 percent. For example, this could reflect the costs of lab tests that could used to determine a person's immune status prior to vaccination. Indeed, numerical investigations suggested unrealistic vaccination schemes when the vaccination term was allowed to have the more standard quadratic form $m = 2$. An alternate way to address the cost differential caused by the density of succeptibles in the population is suggested by Behncke, who introduces a cost function and otherwise maintains a linear functional resulting in a bang-bang control for the vaccination strategy [2].

We assume there are practical limitations on the maximum rate at which individuals who may be vaccinated or treated in a given time period and we define positive constants $\nu_{\max}$ and $\tau_{\max}$ accordingly. We define the set of admissible controls to be

$$\Omega = \{(\nu, \tau) \in L^1(0, T) \mid (\nu(t), \tau(t)) \in [0, \nu_{\max}] \times [0, \tau_{\max}] \; \forall t \in [0, T]\}. \quad (6)$$

We seek an optimal control pair (ν^*, τ^*) such that

$$J(\nu^*, \tau^*) = \min_\Omega J(\nu, \tau). \quad (7)$$

As is listed in Table 1, we assume a maximum vaccination rate of 0.1 and a maximum treatment rate of 0.6. The maximum vaccination rate is based on the 1947 smallpox outbreak in New York City. At the height of this outbreak, a half million to a million of the population of 6 million were vaccinated daily [32]. The maximum treatment

rate was chosen more arbitrarily, but qualitative results were not sensitive to a change of this parameter.

3. **Existence of an optimal control pair.** We begin by examining sufficient conditions for the existence of a solution to the optimal control problem. A high death rate from disease could theoretically cause the population to vanish, and the existence theorem must assume that the external death rate δ is smaller than the population intrinsic growth, or turnover, rate μ.

Theorem 3.1. *There exists an optimal control pair $\nu^*(t)$, $\tau^*(t)$, and corresponding solution S^*, I^*, R^* to the state initial value problem (1) – (4) that minimizes $J(\nu, \tau)$ over Ω.*

Proof. We refer to the conditions in Theorem III.4.1 and its corresponding Corollary in Fleming and Rishel [10]. The requirements there on the set of admissible controls Ω and on the set of end conditions are clearly met here. The following nontrivial requirements from Fleming and Rishel's theorem are listed and later verified below:

A. The set of all solutions to system (1) – (4) with corresponding control functions in Ω (as given in equation (6)) is nonempty.
B. The state system can be written as a linear function of the control variables with coefficients dependent on time and the state variables.
C. The integrand L in equation (5) is convex on Ω and additionally satisfies $L(t, S, I, R, \nu, \tau) \geq c_1 |(\nu, \tau)|^\beta - c_2$, where $c_1 > 0$ and $\beta > 1$.

In order to establish condition A, we refer to Theorem 3.1 by Picard-Lindelöf in Coddington and Levinson [7]. If the solutions to the state equations are *a priori* bounded and if the state equations are continuous and Lipschitz in the state variables, then there is a unique solution corresponding to every admissible control pair in Ω.

Thus, we begin establishing bounds on $N = S + I + R$, and, by extension, S, I, and R. Note that N satisfies the modified logistic equation

$$\frac{dN}{dt} = \mu N \left(1 - \frac{N}{K}\right) - \delta I. \tag{8}$$

and assume that $N(0) = N_0$. If $N > K$, then N is decreasing. Thus, N is bounded above by $\max(N_0, K)$. Note that the quantities S, I, and R decrease only proportional to their present sizes, respectively, and thus none of S, I, and R can be negative. Therefore, the upper bound for N is also an upper bound for S, I, and R.

Note that equations (1) and (2) have division by N, and thus we need a strictly positive lower bound for N. To produce this bound, we note that if the solution to the differential equation

$$\frac{d\tilde{N}}{dt} = \mu \tilde{N} \left(1 - \frac{\tilde{N}}{K}\right) - \delta \tilde{N}. \tag{9}$$

is bounded below then surely the solution to (8) is bounded below by the same bound. Since $\frac{d\tilde{N}}{dt}$ is proportional to $\tilde{N}$ we know that in finite time $\tilde{N}$ remains strictly positive. Thus, for a given end-time T solutions to (8) will have a positive lower bound.

With the bounds established above, it follows that the state system is continuous and bounded. It is equally direct to show the boundedness of the partial derivatives

with respect to the state variables in the state system, which establishes that the system is Lipschitz with respect to the state variables (see [6], page 248). This completes the proof that condition A holds.

Condition B is verified by observing the linear dependence of the state equations on controls ν and τ. Finally, to verify condition C we note that since the integrand L of the objective functional is quadratic in the controls, L is clearly convex in the controls. To prove the bound on the L we note that by the definition of Ω we have $B_2\nu^2 \leq B_2$, and so $B_2\nu^2 - B_2 \leq 0$. Thus,

$$
\begin{aligned}
L &= B_1 I(t) + B_2\left[\frac{R}{K}\right]^m \nu^2(t) + B_3\tau^2(t) \geq B_3\tau^2(t) \geq B_2\nu^2 + B_3\tau^2 - B_2 \\
&\geq \min(B_2, B_3)(\nu^2 + \tau^2) - B_2.
\end{aligned}
$$

$\square$

4. **Characterization of optimal controls.** We apply Pontryagin's Maximum Principle [28] and convert the optimization problem described in the previous section to the problem of finding the point-wise minimum relative to ν and τ of the Hamiltonian.

Theorem 4.1. *Given the optimal control pair (ν^*, τ^*) and corresponding solutions to the state system (1) – (4) S^*, I^*, R^* that minimize the objective functional (5) there exist adjoint variables λ_1, λ_2, and λ_3 satisfying*

$$
\begin{aligned}
\frac{d\lambda_1}{dt} &= \frac{\mu}{K}\left[\lambda_1(S + N - K) + \lambda_2 I + \lambda_3 R\right] \\
&\quad + \beta\frac{I(N-S)}{N^2}(\lambda_1 - \lambda_2) + \nu(\lambda_1 - \lambda_3), \tag{10}
\end{aligned}
$$

$$
\begin{aligned}
\frac{d\lambda_2}{dt} &= -B_1 + \frac{\mu}{K}\left[\lambda_1(S - K) + \lambda_2(N + I) + \lambda_3 R\right] \\
&\quad + \beta\frac{S(N-I)}{N^2}(\lambda_1 - \lambda_2) + \delta\lambda_2 + (\gamma + \tau)(\lambda_2 - \lambda_3), \tag{11}
\end{aligned}
$$

$$
\begin{aligned}
\frac{d\lambda_3}{dt} &= nB_2\nu^2\frac{R^{m-1}}{K^m} + \frac{\mu}{K}\left[(S - K)\lambda_1 + I\lambda_2 + (N + R)\lambda_3\right] \\
&\quad + \beta\frac{SI}{N^2}(\lambda_2 - \lambda_1)), \tag{12}
\end{aligned}
$$

with transversality conditions

$$
\lambda_1(T) = \lambda_2(T) = \lambda_3(T) = 0. \tag{13}
$$

Furthermore, as long as the optimal removed class R^* is nonzero, we may characterize the optimal pair by the continuous functions

$$
\begin{aligned}
\nu^* &= \min\left(\max\left(0, \frac{S^*(\lambda_1 - \lambda_3)}{2B_2\left[\frac{R^*}{K}\right]^m}\right), \nu_{\max}\right), \\
\tau^* &= \min\left(\max\left(0, \frac{I^*(\lambda_2 - \lambda_3)}{2B_3}\right), \tau_{\max}\right). \tag{14}
\end{aligned}
$$

We note that the restriction on R^* can easily be resolved by assuming a nonzero initial value for R_0 for mathematical convenience. Physically this requirement is not significant.

Proof. The result follows from a direct application of a version of Pontryagin's Maximum Principle for bounded controls (see [20, 23, 28]). We form the Hamiltonian H:

$$
\begin{aligned}
H =\ & B_1 I + B_2 \left[\frac{R}{K}\right]^m \nu^2 + B_3 \tau^2 + \lambda_1 \left[\mu N - \mu \frac{N}{K} S - \beta \frac{SI}{N} - \nu S\right] + \\
& \lambda_2 \left[\beta \frac{SI}{N} - (\delta + \gamma + \tau) I - \mu \frac{NI}{K}\right] + \lambda_3 \left[(\gamma + \tau) I + \nu S - \mu \frac{NR}{K}\right]. \ (15)
\end{aligned}
$$

As dictated by the Maximum Principle, the adjoint equations are given by the equations $\frac{d\lambda_1}{dt} = -\frac{\partial H}{\partial S}$, $\frac{d\lambda_2}{dt} = -\frac{\partial H}{\partial I}$, $\frac{d\lambda_3}{dt} = -\frac{\partial H}{\partial R}$, and must satisfy transversality conditions $\lambda_i(T) = 0$ for values $i = 1, 2, 3$. Finally, the optimality conditions dictate that $\frac{\partial H}{\partial \nu} = \frac{\partial H}{\partial \tau} = 0$ for the optimal pair (ν^*, τ^*) on the interior of the control set, and this condition is simplified in equations (14) with attention to the bounds on the pair as given in the definition of Ω in equation (6). Note that as a result of the transversality condition, the optimal vaccination and treatment will be zero at the end time. Observe that by the characterization of the controls given in (14) and the nonzero assumption for R, it follows that the controls are continuous in time. □

The *optimality system* is defined as is the compilation of the state equations (1) – (3), the initial conditions (4), the adjoint equations (10) – (12), and the transversality conditions (13), with the optimality equations (14) substituted into the state and adjoint equations. Uniqueness of the optimality system can be shown for a small time interval.

Theorem 4.2. *For T sufficiently small the optimality system is unique.*

The proof of the the theorem is rather lengthy due to the complexity of the optimality system. The approach is explained in [9, 18], and the proof itself is given in the appendix.

5. **Variations of the SIR model.** As Hethcote details in [13], there are hundreds of variations of the SIR model described in Section 2. Two classic variations result from adding an incubation period before a subject becomes infectious, the SEIR model, or from including waning immunity, the SIRS model. In this section, we provide the system and optimal control differential equations for several variations of the SIR model. We omit the derivations and theory that are similar to the results presented in the previous section for the SIR model. In the section on numerical results we contrast the models and isolate the parameters and terms that cause the most variation in optimal control mitigation strategies.

5.1. **The SEIR model.** Adding an incubation period to the disease creates the "exposed" class E of subjects who have be exposed to the disease and will ultimately become infectious after a waiting period of $\frac{1}{\epsilon}$. An obvious consequence of the addition of a waiting period is the increased window of time to vaccinate subjects before the first big wave of infectious individuals is released into the population. We will examine other qualitative model differences in Section 6. The state equations

for this model are

$$\frac{dS}{dt} = \mu N - \beta\frac{SI}{N} - \nu S - \mu\frac{NS}{K}, \tag{16}$$

$$\frac{dE}{dt} = \beta\frac{SI}{N} - \epsilon E - \mu\frac{NE}{K}, \tag{17}$$

$$\frac{dI}{dt} = \epsilon E - (\gamma + \tau + \delta)I - \mu\frac{NI}{K}, \tag{18}$$

$$\frac{dR}{dt} = (\gamma + \tau)I + \nu S - \mu\frac{NR}{K}. \tag{19}$$

For comparison purposes, we seek to minimize the same objective functional given in equation (5), which we repeat here for convenience:

$$J = \min \in 0^T \left[B_1 I + B_2 \left[\frac{R}{K}\right]^m \nu^2 + B_3\tau^2 \right] dt.$$

Following Theorem 4.1, we consider the Hamiltonian equation given by

$$
\begin{aligned}
H &= B_1 I + B_2 \left[\frac{R}{K}\right]^m \nu^2 + B_3\tau^2 + \\
&\quad \lambda_1 \left[\mu(S + E + I + R) - \beta\frac{SI}{(S + E + I + R)} - \nu S - \right. \\
&\quad \left. \mu\frac{(S + E + I + R)S}{K} \right] + \\
&\quad \lambda_2 \left[\beta\frac{SI}{(S + E + I + R)} - \epsilon E - \mu\frac{(S + E + I + R)E}{K} \right] + \\
&\quad \lambda_3 \left[\epsilon E - (\delta + \gamma + \tau)I - \mu\frac{(S + E + I + R)I}{K} \right] + \\
&\quad \lambda_4 \left[\gamma I + \nu S + \tau I - \mu\frac{(S + E + I + R)R}{K} \right].
\end{aligned}
$$

and the optimal controls are found by solving equations (16) – (19) with initial conditions $S_0, E_0, I_0,$ and R_0 along with the adjoint equations

$$
\begin{aligned}
\frac{d\lambda_1}{dt} &= -\lambda_1\mu + (\lambda_1 - \lambda_2)\beta I\frac{N - S}{N^2} + (\lambda_1 - \lambda_4)\nu \\
&\quad + \frac{\mu}{K}\left[(S + N)\lambda_1 + E\lambda_2 + I\lambda_3 + R\lambda_4 \right] \tag{20}
\end{aligned}
$$

$$
\begin{aligned}
\frac{d\lambda_2}{dt} &= -\lambda_1\mu + (\lambda_2 - \lambda_1)\beta\frac{SI}{N^2} + (\lambda_2 - \lambda_3)\epsilon \\
&\quad + \frac{\mu}{K}\left[S\lambda_1 + (E + N)\lambda_2 + I\lambda_3 + R\lambda_4 \right] \tag{21}
\end{aligned}
$$

$$
\begin{aligned}
\frac{d\lambda_3}{dt} &= -B_1 - \lambda_1\mu + (\lambda_1 - \lambda_2)\beta S\frac{N - I}{N^2} + \delta\lambda_3 \\
&\quad + (\lambda_3 - \lambda_4)(\gamma + \tau) + \frac{\mu}{K}\left[S\lambda_1 + E\lambda_2 + (I + N)\lambda_3 + R\lambda_4 \right] \tag{22}
\end{aligned}
$$

$$
\begin{aligned}
\frac{d\lambda_4}{dt} &= nB_2\nu^2\frac{R^{m-1}}{K^m} - \lambda_1\mu + (\lambda_2 - \lambda_1)\beta\frac{SI}{N^2} + \\
&\quad \frac{\mu}{K}\left[\lambda_1 S + \lambda_2 E + \lambda_3 I + \lambda_4(R + N) \right], \tag{23}
\end{aligned}
$$

and transversality conditions as given in equation (13) of Theorem 4.1. The optimal
control pair satisfies

$$\nu^* = \min\left(\max\left(0, \frac{S^*(\lambda_1 - \lambda_4)}{2B_2\left[\frac{R^*}{K}\right]^m}\right), \nu_{\max}\right),\tag{24}$$

$$\tau^* = \min\left(\max\left(0, \frac{I^*(\lambda_3 - \lambda_4)}{2B_3}\right), \tau_{\max}\right).\tag{25}$$

$$\tag{26}$$

5.2. **The SIRS model.** We now consider the consequences when the SIR model
is modified to account for waning immunity. We assume in this model that after a
time period $\frac{1}{\omega}$, subjects in the removed class return to the susceptible class. Again,
one consequence of this variation is intuitively clear. The susceptible class will
likely have a long term positive equilibrium extending the need for vaccination. We
discuss this and other significant outcomes in the consideration of this model in
Section 6.

The state equations for the SIRS model are

$$\frac{dS}{dt} = \mu N - \beta\frac{SI}{N} - \nu S - \mu\frac{NS}{K} + \omega R,\tag{27}$$

$$\frac{dI}{dt} = \beta\frac{SI}{N} - (\gamma + \tau + \delta)I - \mu\frac{NI}{K},\tag{28}$$

$$\frac{dR}{dt} = (\gamma + \tau)I + \nu S - \mu\frac{NR}{K} - \omega R.\tag{29}$$

The objective functional is as in equation (5) of Section 2. The Hamiltonian is
formed as before, resulting in the adjoint equations

$$\frac{d\lambda_1}{dt} = \frac{\mu}{K}\left[\lambda_1(S + N - K) + \lambda_2 I + \lambda_3 R\right] +$$
$$\beta\frac{I(N - S)}{N^2}(\lambda_1 - \lambda_2) + \nu(\lambda_1 - \lambda_3)\tag{30}$$

$$\frac{d\lambda_2}{dt} = -B_1 + \frac{\mu}{K}\left[\lambda_1(S - K) + \lambda_2(N + I) + \lambda_3 R\right]$$
$$+ \beta\frac{S(N - I)}{N^2}(\lambda_1 - \lambda_2) + \delta\lambda_2 + (\gamma + \tau)(\lambda_2 - \lambda_3)\tag{31}$$

$$\frac{d\lambda_3}{dt} = nB_2\nu^2\frac{R^{m-1}}{K^m} + \frac{\mu}{K}\left[(S - K)\lambda_1 + I\lambda_2 + (N + R)\lambda_3\right]$$
$$+ \beta\frac{SI}{N^2}(\lambda_2 - \lambda_1)) - \omega(\lambda_1 - \lambda_3).\tag{32}$$

The transversality conditions are as in equation (13) of Section 2, and the optimal
control pair will satisfy the same condition given in equations (14) in Theorem 4.1.

6. **Numerical results.** The optimal control problem was solved using an iterative
scheme derived by Hackbush [11] and described briefly here. Using an initial guess
for the control variables, ν and τ, the state variables, S, E (if applicable), I, R, are
solved forward in time, and then the adjoint variables, λ_i, are solved backwards in
time. If the new values of the state and adjoint variables differ from the previous
values, the new values are used to update ν and τ and the process is repeated until
the system converges. For the SIR model, the state equations are given by equations
(1) – (3) with the initial conditions given by (4). The adjoint equations are given

by equations (10) – (12) with the final time conditions given by (13). Finally, the control variables are given by equations (14). Similarly, the SEIR model has state equations (16) – (19), adjoint equations (20) – (23), and control variables (24) – (25). Finally, the SIRS model has state equations (27) – (29), adjoint equations (30) – (32), and control variables (13). For each run, the control variables are set to an initial guess of 0.0 for all time. In the numerical runs presented here we use the parameter assumptions listed in Table 1.

In sensitivity analyses, we found that the weights in the objective functional (5) had very little qualitative impact on the suggested vaccination scheme, though as we mentioned in Section 2 the inclusion of the term $(\frac{R}{K})^{10}$ in the functional was essential to the outcome. Likewise, the carrying capacity K can be varied widely provided that it is not extremely small with respect to the number of initially infectious individuals. The incubation period ϵ specified in the SEIR model does affect the number of infectious individuals in the population at any time – the longer the incubation period the more infectives – but the dependence is smooth and the qualitative behaviors the same. Likewise, varying the waning rate ω has the expected impact in the SIRS model – the higher the waning rate, the higher the number of newly susceptibles, and thus the higher the resulting maintenance vaccination level. On the other hand, the models are sensitive to changes in the incidence rate β and the death rate δ.

In our numerical runs, we compare optimal mitigation strategies as we vary incidence rates, death rates, and weights for the relative costs of vaccination and treatment. With a death rate of 0 we varied the incidence rates from 0.05 to 0.55, and for an incidence rate of 0.3 we varied the death rate from 0 to 0.1. Recalling that a factor of 1000 is needed to balance units for treatment and vaccination with the number of infected individuals in the objective functional, we shifted a total weight of 2000 between treatment and vaccination costs. To investigate the required effort for the predicted optimal control, we multiplied the controls by the corresponding states. For treatment, we multiplied the treatment rate by the number of infected individuals in the population at that time to give us an estimate for how much effort this represented. For vaccination, we multiplied the vaccination rate by the number of susceptible individuals in the population for a similar estimate of effort. Our graphs show by color the number of individuals who are infected, treated, or vaccinated in any time step for the various scenarios.

The advice suggested by the various runs is consistent, as is detailed below and illustrated in Figures 1 through 4. Regardless of the relative costs of vaccination and treatment and also regardless of the disease structure, it is essential for the effective control of an epidemic to vaccinate at the highest possible rate as early as possible to minimize the number of individuals who become infectious and to minimize the effort of controlling the epidemic. However, if the relative costs of treatment and vaccination allow treatment to coincide with vaccination, we see that the presence of treatment reduces the size of the infected population correspondingly. While it is not evident in the time scale pictured for this article, we further observe in the SIRS case the need for continued vaccination for small portions of the large susceptible population, as one would expect to see where there is waning immunity.

In Figures 1 and 2 we see the baseline recommendations for vaccination and treatment and the resulting numbers of infected when vaccination and treatment are equally weighted and incident rates and disease death rates vary. Note that the true maximum numbers who are vaccinated or treated at any one time are about
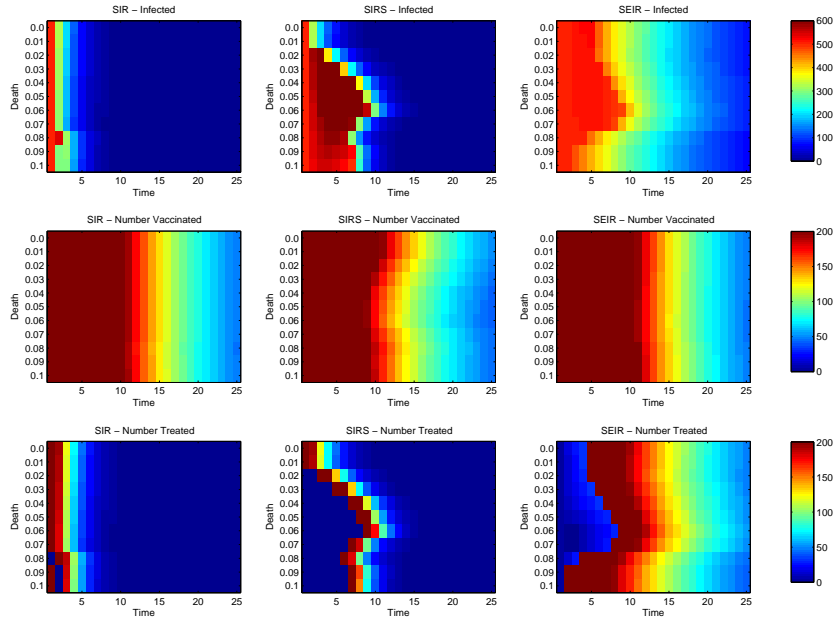
FIGURE 1. For this figure, vaccination and treatment are weighted equally, the incidence rate is $\beta = 0.3$, and the death rate δ varies. All plots show the solution of the model systems using the predicted optimal control solutions for each given value of δ. The graphs in the first column are the results of the SIR model, the second column the SIRS model, and the third column the SEIR model. The first row shows the size of the infected class versus time for each disease-related death rate for each model, the second row shows the number vaccinated, and the third shows the number treated. It is important to note that the color scale for the infected class is 0-600, while the scale for the vaccinated and treated is only 0-200. For example, we see that when $\delta = 0$, which is the top-most bar in each plot, there is little variation between the SIR and SIRS model for the size of the infected class. However, as the disease-related death rate increases to 0.1, there is a noted increase in the size of the infected class for SIRS and SEIR models that is not reflected in the SIR model. The optimal number of vaccinations does not vary widely from model to model for varying δ. The timing of treatment does vary from model to model and across values for δ.

550 and 400, respectively, and our smaller scale in the graph is to visually aid with comparisons as time progresses. One can observe in these initial baseline cases that treatment is sometimes foregone to support the necessary immunization; however, there is a corresponding price in the number of individuals that become infected. One can also observe the continuing vaccination in the SIRS cases until the disease is eradicated.
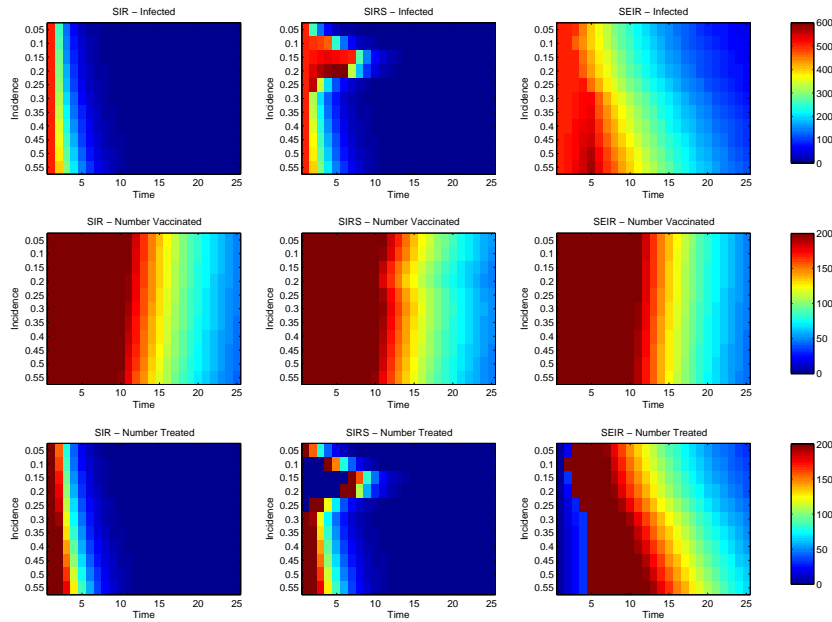
FIGURE 2. For this figure, vaccination and treatment are weighted equally, the incidence rate β varies, and the death rate δ is zero. All plots show the solution of the model systems using the predicted optimal control solutions for each given value of β. The graphs in the first column are the results of the SIR model, the second column the SIRS model, and the third column the SEIR model. The first row shows the size of the infected class versus time for each disease-related death rate for each model, the second row shows the number vaccinated, and the third shows the number treated. It is important to note that the color scale for the infected class is 0-600, while the scale for the vaccinated and treated is only 0-200. Similar to the results of Figure 1, the optimal solution shows maximum vaccination across all models and parameter values for at least the first half of the run. Treatment levels, however, vary by model and the incidence values, β.

Figure 3 considers the case in which vaccination carries a weight of 200 and treatment carries a weight of 1800, so that treatment is 9 times as expensive as vaccination. This figure should be compared to Figure 2 in which treatment and vaccination are equally weighted. We see that the SIRS population benefits from the less expensive vaccination (observe the missing "bumps" in the second column of Figure 3) because some of the resources saved in the inexpensive vaccination effort can be applied to the necessary treatment. In contrast, the SEIR population suffers because there are not enough resources to treat at the higher rate used in the baseline case for higher incidence rates. The graph shown is for varying incidence, but the results are similar as the death rate varies. The results are also qualitatively
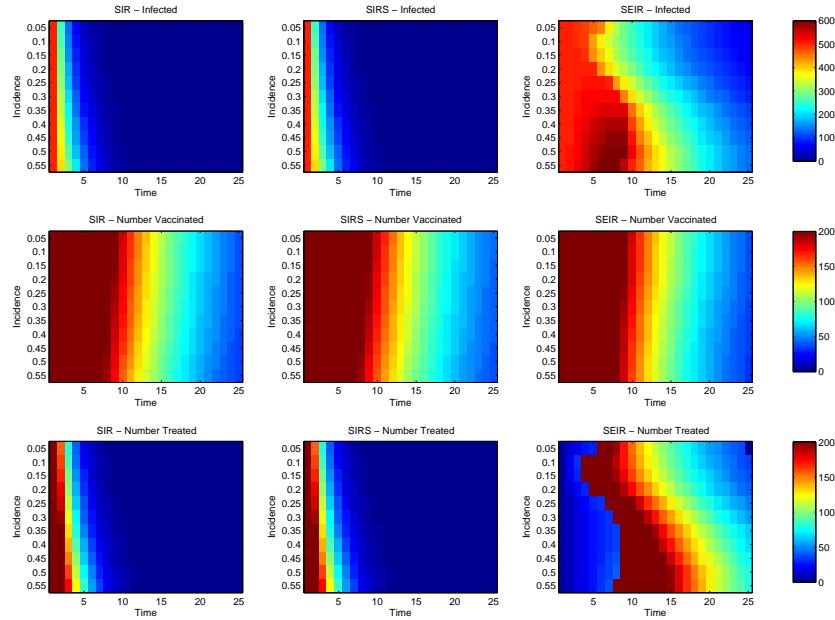
FIGURE 3. For this figure, treatment was nine times more expensive than vaccination (vaccination weight = 200, treatment weight = 1800), the incidence rate β varies, and the death rate δ is zero. All plots show the solution of the model systems using the predicted optimal control solutions for each given value of β. The graphs in the first column are the results of the SIR model, the second column the SIRS model, and the third column the SEIR model. The first row shows the size of the infected class versus time for each disease-related death rate for each model, the second row shows the number vaccinated, and the third shows the number treated. It is important to note that the color scale for the infected class is 0-600, while the scale for the vaccinated and treated is only 0-200. Note the increased size of the infected class for the SEIR model as compared with that in Figure 2 as a function of increased cost for treatment.

the same when treatment is only 3 times as expensive or as much as 200 times more expensive.

Finally, Figure 4 shows the results as we vary incidence when vaccination is 9 times as expensive as treatment. Again, this figure should be compared to Figure 2 in which treatment and vaccination have equal cost. Ironically, even though treatment is relatively cheap we have a big hole in the numbers treated for the SIRS case, reflecting the urgency of vaccination, and yet the lack of treatment produces a surge of infected individuals. As we would expect from the previous example, the SEIR audience fairs better because the less expensive treatment has a
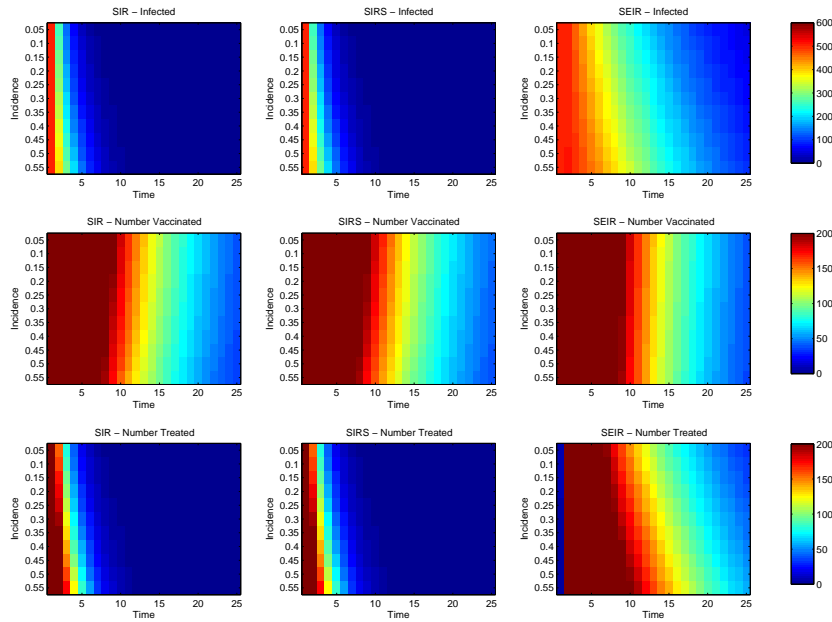
FIGURE 4. For this figure, vaccination was nine times more expensive than treatment (vaccination weight = 200, treatment weight = 1800), the incidence rate β varies, and the death rate δ is zero. All plots show the solution of the model systems using the predicted optimal control solutions for each given value of β. The graphs in the first column are the results of the SIR model, the second column the SIRS model, and the third column the SEIR model. The first row shows the size of the infected class versus time for each disease-related death rate for each model, the second row shows the number vaccinated, and the third shows the number treated. It is important to note that the color scale for the infected class is 0-600, while the scale for the vaccinated and treated is only 0-200. Again, the SEIR infected class shows the most dramatic impact of the cost of treatment in comparison to Figure 2.

strong effect on the number who become infected when there is an exposed period for the disease.

Finally, Figure 4 shows the results as we vary incidence when vaccination is 9 times as expensive as treatment. Again, this figure should be compared to Figure 2 in which treatment and vaccination have equal cost. Ironically, even though treatment is relatively cheap we have a big hole in the numbers treated for the SIRS case, reflecting the urgency of vaccination, and yet the lack of treatment produces a surge of infected individuals. As we would expect from the previous example, the SEIR audience fairs better because the less expensive treatment has a strong effect on the number who become infected when there is an exposed period for the disease.

Another view of the mitigation strategies suggested by the optimal control simulations can be seen in Figure 5. In this figure, we observe the number of individuals vaccinated and treated on top, and the rates of vaccination and treatment on the bottom. We see that it can be misleading to only record only the recommended rates because at the end of the time period the numbers of infected are so low that a high treatment rate corresponds to a small number of infected individuals being treated. We note that the strategy requires vaccinating at the maximum rate allowed, and when that translates to fewer succeptible individuals remaining and being vaccinated, then additional healthcare dollars can be devoted to treatment while maintaining the recommended vaccination rate. This time-series view of a given scenario helps to illustrate the optimal control recommendation of maximum vaccination as a primary strategy.
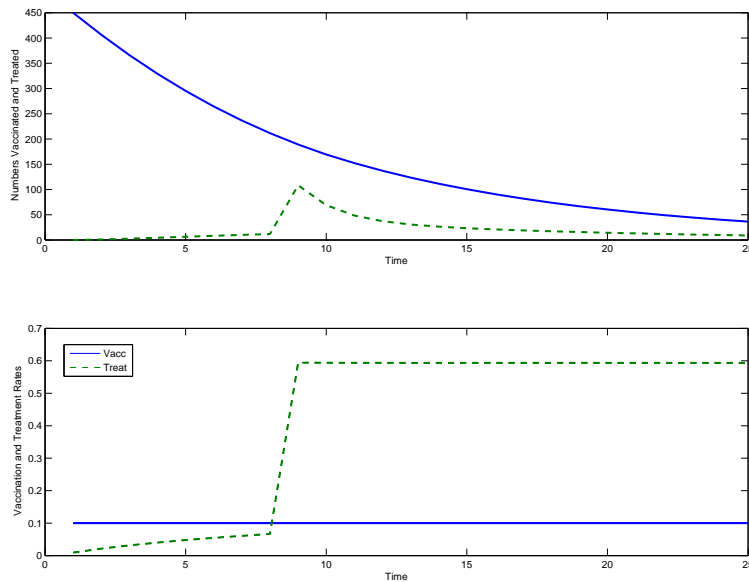

FIGURE 5. For this figure, we have selected the SEIR model run from the scenario shown in Figure 3 where the vaccination weight = 200, treatment weight = 1800, incidence rate =0.35 and the death rate as zero. The top time series shows the predicted numbers of individuals vaccinated (solid line) and treated (dotted line) in the first 25 time steps. In contrast the lower time series shows the predicted optimal control rates for vaccination (solid) and treatment (dotted). These plots show that a prediction of a constant vaccination or treatment rate does not imply a constant level of effort, but rather the number of individuals affected provides a better view of the true costs.

7. **Conclusion.** While underlying epidemic structure is crucial to answering many questions, it appears that for the range of models we studied there is only subtle

variation in the optimal strategies. Our conclusion from our simulations is that vaccination, if available, is an essential tool in fighting an epidemic. However, if resources allow for the provision of treatment as well, this additional tool is a valuable resource in decreasing the number of individuals who are affected by an epidemic, particularly if a disease has an exposed period and/or a high incidence rate.

REFERENCES

[1] R. M. Anderson, *Population ecology of infectious disease agents*, "Theoretical Ecology, Principles, and Applications" (ed. R. M. May), Blackwell, (1981), 318–355.

[2] Horst Behncke, *Optimal control of deterministic epidemics*, Optimal Control Applied Methods, **21** (2000), 269–285.

[3] D. Bernoulli and S. Blower, *An attempt at a new analysis of the mortality caused by smallpox and the advantages of inoculation to prevent it*, Reviews in Medical Virology, **14** (2004), 275–288.

[4] C. Bowman, A. B. Gumel, P. van den Driessche, J. Wu and H. Zhu, *A mathematical model for assessing control strategies against west nile virus*, Bulletin of Mathematical Biology, **67** (2005), 1107–1133.

[5] Fred Brauer, *Some simple epidemic models*, Math. BioSci. Eng., **3** (2006), 1–15.

[6] Earl A. Coddington, "An Introduction to Ordinary Differential Equations," Prentice-Hall Inc, 1961, 292.

[7] Earl A. Coddington and Norman Levinson, "Theory of Ordinary Differential Equations," McGraw Hill Co. Inc., 1955, 429.

[8] Neil M. Ferguson, Matt J. Keeling, W. John Edmunds, Raymond Gani, Bryan T. Grenfell, Roy M. Anderson and Steve Leach, *Planning for smallpox outbreaks*, Nature, **425** (2003), 681–685.

[9] K. R. Fister, S. Lenhart and J. McNally, *Optimizing chemotherapy in an HIV*, Elec Journal of DE, **1998** (1998), 1–12.

[10] Wendell H. Fleming and Raymond W. Rishel, "Deterministic and Stochastic Optimal Control," Applications of Mathematics, No. 1, Springer-Verlag, 1975.

[11] W. K. Hackbush, *A numerical method for solving parabolic equations with opposite orientations*, Computing, **20** (1978), 229–240.

[12] Herbert Hethcote, Ma Zhien and Liao Shengbing, *Effect of quarantine in six endemic models for infectious diseases*, Mathematical Biosciences, **180** (2002), 141–160.

[13] H. W. Hethcote, *A thousand and one epidemic models*, "Frontiers in Theoretical Biology" (ed. S. A. Levin), Springer-Verlag, (1994) 504–515.

[14] H. W. Hethcote and P. van den Driessche, *Some epidemiological models with nonlinear incidence*, Journal of Mathematical Biology, **29** (1991), 271–287.

[15] Herbert W. Hethcote, *Optimal ages of vaccination for measles*, Mathematical Biosciences, **89** (1988), 29–52.

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

[17] H. W. Hethcote, H. W. Stech and P. van den Driessche, *Periodicity and stability in epidemic models: A survey*, "Differential Equations and Applications in Ecology, Epidemics, and Population Problems," Academic Press (1981), 65–82.

[18] Hem Raj Joshi, *Optimal control of an HIV immunology model*, Optimal Control Applications & Methods, **23** (2002), 199–213.

[19] E. Jung, S. Lenhart and Z. Feng, *Optimal control of treatments in a two-strain tuberculosis*, Discrete and Continuous Dynamical Systems. Series B, **2** (2002), 473–482.

[20] Morton I. Kamien and Nancy L. Schwartz, "Dynamic Optimization," "Advanced Textbooks in Economics V. **31**," North-Holland Publishing Co., 1991.

[21] Edward H. Kaplan, David L. Craft and Lawrence M. Wein, *Emergency response to a smallpox attack: The case for mass vaccination*, Proceedings of the National Academy of Sciences of the United States of America, **99** (2002), 10935–10940.

[22] Urszula Ledzewicz and Heinz Schaettler, *On optimal controls for a general mathematical model for chemotherapy of HIV*, Proceedings of the 2002 American Control Conference, Anchorage, Alaska, (2002), 3454–3459.

[23] Suzanne L. Lenhart and John T. Workman, "Optimal Control Applied to Biological Models," Chapman Hall/CRC, 2007.

[24] Marc Lipsitch, Ted Cohen, Ben Cooper, James M. Robins, Stefan Ma, Lyn Jamesi, Gowri Gopalakrishna, Suok Kai Chew, Chorh Chuan Tan, Matthew H. Samore, David Fisman and Megan Murray, *Transmission dynamics and control of severe acute respiratory syndrome*, Science, **300** (2003), 1966–1970.

[25] Wei Min Liu, Herbert W. Hethcote and Simon A. Levin, *Dynamical behavior of epidemiological models with nonlinear incidence rates*, Journal of Mathematical Biology, **25** (1987), 359–380.

[26] Wei Min Liu, Simon A. Levin and Yoh Iwasa *Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models*, Journal of Mathematical Biology, **23** (1986), 187–204.

[27] C. Marchetti, P. S. Meyer and J. H. Ausubel, *Human population dynamics revisited with the logistic model: How much can be modeled and predicted?* Human Evolution, **19** (2004), 247–285.

[28] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mishchenko, "The Mathematical Theory of Optimal Processes," Interscience Publishers John Wiley & Sons, Inc. New York-London, 1962, 360.

[29] Steven Riley, Christophe Fraser, Christl A. Donnelly , Azra C. Ghani, Laith J. Abu-Raddad, Anthony J. Hedley, Gabriel M. Leung, Lai-Ming Ho, Tai-Hing Lam, Thuan Q. Thach, Patsy Chau, King-Pan Chan, Su-Vui Lo, Pak-Yin Leung, Thomas Tsang, William Ho, Koon-Hung Lee, Edith M. C. Lau, Neil M. Ferguson and Roy M. Anderson, *Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions*, Science, **300** (2003), 1961–1966.

[30] S. P. Sethi, *Quantitative Guidelines for communicable disease control program: A complete synthesis*, Biometrics, **30** (1974), 681–691.

[31] Felippe de Souza, M. A. L. Caetano and Takashi Yonoyama, *Optimal control theory applied to the anti-viral treatment of AIDS*, Decision and Control, 2000, Proceedings of the 39th IEEE Conference V. **5**, IEEE, (2000), 4839–4844.

[32] Lorna Thorpe, Farzad Mostashari, Adam Karpati, Steven Schwartz, Susan Manning, Malissa Marx and Thomas Frieden, *Mass smallpox vaccination and cardiac deaths, New York City, 1947*, Emerging Infectious Disease, **10** (2004).

[33] Wendi Wang and Shigui Ruan, *Simulating the SARS outbreak in Beijing with limited data*, Journal of Theoretical Biology, **227** (2004), 369–379.

Appendix. We include the proof of Theorem 4.2.

Proof. In order to consider the uniqueness of the optimal control as characterized above, we must first consider bounds on the adjoint system. Note that bounds for the state system were established in Theorem 3.1. To see that the adjoint system is bounded, we rearrange equations (10) – (12):

$$
\begin{aligned}
\frac{d\lambda_1}{dt} &= \lambda_1\left[\frac{\mu}{K}(S+N-K)+\beta\frac{I(N-S)}{N^2}+\nu\right] \\
&\quad + \lambda_2\left[\frac{\mu}{K}I-\beta\frac{I(N-S)}{N^2}\right]+\lambda_3\left[\frac{\mu}{K}R-\nu\right] \\
\frac{d\lambda_2}{dt} &= -B_1+\lambda_1\left[\frac{\mu}{K}(S-K)+\beta\frac{S(N-I)}{N^2}\right] \\
&\quad + \lambda_2\left[\frac{\mu}{K}(N+I)-\beta\frac{S(N-I)}{N^2}+\delta+\gamma+\tau\right] \\
&\quad + \lambda_3\left[\frac{\mu}{K}R-\gamma-\tau\right]
\end{aligned}
$$

$$\frac{d\lambda_3}{dt} = mB_2\nu^2\frac{R^{m-1}}{K^m} + \lambda_1\left[\frac{\mu}{K}(S-K) - \beta\frac{SI}{N^2}\right]$$
$$+\lambda_2\left[\frac{\mu}{K}I + \beta\frac{SI}{N^2}\right] + \lambda_3\left[\frac{\mu}{K}(N+R)\right].$$

We recall that $N = S + I + R$, that S, I, and R are nonnegative, and that $N > 0$. Thus, $0 \le \frac{I(N-S)}{N^2}, \frac{IS}{N^2}, \frac{S(N-I)}{N^2} \le 1$. We recall additionally the bounds $0 \le \nu \le \nu_{\max}$ and $0 \le \tau \le \tau_{\max}$, and with this substitution we see that the adjoint system is bounded by linear systems with bounded coefficients. Thus, the sub- and super-solutions are uniformly bounded, establishing bounds for the adjoint system in finite time.

Now, suppose that there are two solutions to the optimality system:

$$(S(t), I(t), R(t), \lambda_1(t), \lambda_2(t), \lambda_3(t)) \text{ and } (\bar{S}(t), \bar{I}(t), \bar{R}(t), \bar{\lambda}_1(t), \bar{\lambda}_2(t), \bar{\lambda}_3(t)).$$

To show that the two solutions are equivalent, it is convenient to make a change of variables. We define s, i, r, ϕ_1, ϕ_2, ϕ_3, $\bar{s}$, $\bar{i}$, $\bar{r}$, $\bar{\phi}_1$, $\bar{\phi}_2$, and $\bar{\phi}_3$ so that

$$S(t) = e^{\alpha t}s(t),\ I(t) = e^{\alpha t}i(t),\ R(t) = e^{\alpha t}r(t),$$

$$\lambda_1(t) = e^{-\alpha t}\phi_1(t),\ \lambda_2(t) = e^{-\alpha t}\phi_2(t),\ \lambda_3(t) = e^{-\alpha t}\phi_3(t),$$

where α is a constant to be chosen later. The barred variables are transformed similarly. Note that the bounds for the state and adjoint variables can be extended to bounds for the new variables. With the new variables the optimality conditions become

$$\nu = \min\left(\max\left(0, \frac{e^{2\alpha t}s(t)(\phi_1(t) - \phi_3(t))}{2B_2\left[\frac{e^{\alpha t}r(t)}{K}\right]^m}\right), \nu_{\max}\right), \tag{33}$$

$$\tau = \min\left(\max\left(0, \frac{e^{2\alpha t}i(t)(\phi_2(t) - \phi_3(t))}{2B_3}\right), \tau_{\max}\right), \tag{34}$$

and likewise the optimality conditions for the barred equations could be defined. For convenience we define $n(t) = s(t) + i(t) + r(t)$ and we note that $N(t) = e^{\alpha t}n(t)$.

We make the change of variables for each of the six differential equations in our optimality system. For ease of notaiton, we generally omit the dependence on time in the following except in the case taht a specific time is intended. We consider the difference of the resulting equations for s and $\bar{s}$, i and $\bar{i}$, and so on, and we then simplify the resulting equations by integration with the use of appropriate integrating factors.

We consider the first state equation with both solution sets and the change of variables:

$$\alpha e^{\alpha t}s + e^{\alpha t}\dot{s} = \mu e^{\alpha t}n - \beta\frac{e^{2\alpha t}si}{e^{\alpha t}n} - \nu e^{\alpha t}s - \mu\frac{e^{2\alpha t}ns}{K}$$

$$\alpha s + \dot{s} = \mu n - \beta\frac{si}{n} - \nu s - \mu\frac{e^{\alpha t}ns}{K}$$

We subtract from the above the corresponding barred equation and find:

$$\alpha(s - \bar{s}) + (\dot{s} - \dot{\bar{s}}) = \mu(n - \bar{n}) - \beta\left[\frac{si}{n} - \frac{\bar{s}\bar{i}}{\bar{n}}\right]$$
$$-\nu(s - \bar{s}) - \mu e^{\alpha t}\left[\frac{ns}{K} - \frac{\bar{n}\bar{s}}{K}\right].$$

We multiply by $(s - \bar{s})$ and integrate from 0 (at which the state equation variables are equivalent) to T yielding

$$\frac{1}{2}(s(T) - \bar{s}(T))^2 + \alpha \int_0^T (s - \bar{s})^2 \, dt$$

$$= \mu \int_0^T (n - \bar{n})(s - \bar{s}) \, dt - \beta \int_0^T (s - \bar{s}) \left[\frac{si}{n} - \frac{\bar{s}\bar{i}}{\bar{n}} \right] dt$$

$$- \int_0^T (\nu s - \bar{\nu}\bar{s})(s - \bar{s}) \, dt - \frac{\mu}{K} \int_0^T e^{\alpha t}(s - \bar{s}) \left[ns - \bar{n}\bar{s} \right] dt. \qquad (35)$$

The remaining equations are manipulated similarly, with derivations and dependence on time omitted in the interest of space. The specific characterization of the controls as given in equations (33) and (34) is represented simply by ν, $\bar{\nu}$, τ, and $\bar{\tau}$ in the six manipulated equations:

$$\frac{1}{2}(i(T) - \bar{i}(T))^2 + \alpha \int_0^T (i - \bar{i})^2 \, dt = \beta \int_0^T (i - \bar{i}) \left[\frac{si}{n} - \frac{\bar{s}\bar{i}}{\bar{n}} \right] dt$$

$$-(\gamma + \delta) \int_0^T (i - \bar{i})^2 \, dt - \int_0^T (\tau i - \bar{\tau}\bar{i})(i - \bar{i}) \, dt$$

$$-\frac{\mu}{K} \int_0^T e^{\alpha t} \left[ni - \bar{n}\bar{i} \right] (i - \bar{i}) \, dt, \qquad (36)$$

$$\frac{1}{2}(r(T) - \bar{r}(T))^2 + \alpha \int_0^T (r - \bar{r}^2 \, dt = \gamma \int_0^T (i - \bar{i})(r - \bar{r}) \, dt$$

$$-\frac{\mu}{K} \int_0^T e^{\alpha t} + \int_0^T (\tau i - \bar{\tau}\bar{i})(r - \bar{r}) \, dt + \int_0^T (\nu s - \bar{\nu}\bar{s})(r - \bar{r} \, dt$$

$$-\frac{\mu}{K} \int_0^T e^{\alpha t} \left[nr - \bar{n}\bar{r} \right] (r - \bar{r}) \, dt, \qquad (37)$$

$$-\alpha \int_0^T (\phi_1 - \bar{\phi}_1)^2 \, dt - \frac{1}{2}(\phi_1(0) - \bar{\phi}_1(0))^2 = -\mu \int_0^T (\phi_1 - \bar{\phi}_1^2) \, dt$$

$$+\frac{\mu}{K} \int_0^T e^{\alpha t}(\phi_1 - \bar{\phi}_1) \left\{ [\phi_1(s + n) + \phi_2 i + \phi_3 r] \right.$$

$$\left. - \left[\bar{\phi}_1(\bar{s} + \bar{n}) + \bar{\phi}_2\bar{i} + \bar{\phi}_3\bar{r} \right] \right\} dt$$

$$+\beta \int_0^T e^{\alpha t}(\phi_1 - \bar{\phi}_1) \left[\frac{i(n - s)}{n^2}(\phi_1 - \phi_2) - \frac{\bar{i}(\bar{n} - \bar{s})}{\bar{n}^2}(\bar{\phi}_1 - \bar{\phi}_2) \right] dt$$

$$+\int_0^T (\phi_1 - \bar{\phi}_1) \left(\nu(\phi_1 - \phi_3) - \bar{\nu}(\bar{\phi}_1 - \bar{\phi}_3) \right) dt, \qquad (38)$$

$$-\alpha \int_0^T (\phi_2 - \bar{\phi}_2)^2 \, dt - \frac{1}{2}(\phi_2(0) - \bar{\phi}_2(0))^2$$

$$= \frac{\mu}{K} \int_0^T e^{\alpha t}(\phi_1 s - \bar{\phi}_2 \bar{s})(\phi_2 - \bar{\phi}_2) \, dt$$

$$-\mu \int_0^T (\phi_1 - \bar{\phi}_1)(\phi_2 - \bar{\phi}_2) \, dt$$

$$+\frac{\mu}{K} \int_0^T e^{\alpha t} \left[\left(\phi_2(n+i) - \bar{\phi}_2(\bar{n} + \bar{i}) \right) + \left(\phi_3 r - \bar{\phi}_3 \bar{r} \right) \right] (\phi_2 - \bar{\phi}_2) \, dt$$

$$+\beta \int_0^T \left[\frac{s(n-i)}{n^2}(\phi_1 - \phi_2) - \frac{\bar{s}(\bar{n} - \bar{i})}{\bar{n}^2}(\bar{\phi}_1 - \bar{\phi}_2) \right] (\phi_2 - \bar{\phi}_2) \, dt$$

$$+\delta \int_0^T (\phi_2 - \bar{\phi}_2)^2 \, dt$$

$$+\gamma \int_0^T \left[(\phi_2 - \phi_3) - (\bar{\phi}_2 - \bar{\phi}_3) \right] (\phi_2 - \bar{\phi}_2) \, dt$$

$$+ \int_0^T \left[\tau(\phi_2 - \phi_3) - \bar{\tau}(\bar{\phi}_2 - \bar{\phi}_3 \right] (\phi_2 - \bar{\phi}_2) \, dt, \tag{39}$$

$$-\alpha \int_0^T (\phi_3 - \bar{\phi}_3)^2 \, dt - \frac{1}{2}(\phi_3(0) - \bar{\phi}_3(0))^2$$

$$= \frac{mB_2}{K^m} \int_0^T e^{\alpha m t} \left[\nu^2 r^{n-1} - \bar{\nu}^2 \bar{r}^{n-1} \right] (\phi_3 - \bar{\phi}_3) \, dt$$

$$+\frac{\mu}{K} \int_0^T e^{\alpha t}(s\phi_1) - \bar{s}\bar{\phi}_1)(\phi_3 - \bar{\phi}_3) \, dt$$

$$-\mu \int_0^T (\phi_1 - \bar{\phi}_1)(\phi_3 - \bar{\phi}_3) \, dt$$

$$+\frac{\mu}{K} \int_0^T e^{\alpha t} \{ [i\phi_2 + (n+r)\phi_3]$$

$$- \left[i\bar{\phi}_2 + (\bar{n} + \bar{r})\bar{\phi}_3 \right] \} (\phi_3 - \bar{\phi}_3) \, dt$$

$$+\beta \int_0^T \left[\frac{si}{n^2}(\phi_2 - \phi_1) - \frac{\bar{s}\bar{i}}{\bar{n}^2}(\bar{\phi}_2 - \bar{\phi}_1) \right] (\phi_3 - \bar{\phi}_3) \, dt. \tag{40}$$

For convenience we define

$$\Psi(t) = (s(t) - s\bar{(t)})^2 + (i(t) - i\bar{(t)})^2 + (r(t) - r\bar{(t)})^2$$

and

$$\Phi(t) = (\phi_1(t) - \phi_1\bar{(t)})^2 + (\phi_2(t) - \phi_2\bar{(t)})^2 + (\phi_3(t) - \phi_3\bar{(t)})^2.$$

Observe that $\Psi(t) \geq 0$ and $\Phi(t) \geq 0$ for all t.

We multiply equations (38) – (40) by -1 and then add the resulting equations to equations (35) – (37). The left-hand side of the resulting equation becomes

$$\frac{1}{2}\Psi(t) + \frac{1}{2}\Phi(t) + +\alpha \int_0^T [\Psi(t) + \Phi(t)] \, dt,$$

and we work now to bound the right-hand side.

An elementary inequality

$$(x - y)^2 \geq 0 \Rightarrow x^2 + y^2 \geq \frac{1}{2}(x^2 + y^2) \geq xy$$

can be used repeatedly to simplify right-hand expressions. The first instance is in the common form

$$(x - \bar{x})(y - \bar{y}) \leq (x - \bar{x})^2 + (y - \bar{y})^2. \tag{41}$$

Another common expression and needed inequality:

$$
\begin{aligned}
(xy - \bar{x}\bar{y})(w - \bar{w}) &= (xy - \bar{x}y + \bar{x}y - \bar{x}\bar{y})(w - \bar{w}) \\
&= y(x - \bar{x})(w - \bar{w}) + \bar{x}(y - \bar{y})(w - \bar{w}) \\
&\leq y^2(x - \bar{x})^2 + \bar{x}^2(y - \bar{y})^2 + 2(w - \bar{w})^2 \\
&\leq C\left[(x - \bar{x})^2 + (y - \bar{y})^2 + (w - \bar{w})^2\right] \tag{42}
\end{aligned}
$$

where C depends on bounds for $\bar{x}$ and y.

Several bounds require in turn a bound on $(n - \bar{n})^2$ which follows directly from the definition of n:

$$(n - \bar{n})^2 \leq (s - \bar{s})^2 + (i - \bar{i})^2 + (r - \bar{r})^2. \tag{43}$$

A bound is needed for the expressions with division by n^2 in equations (38) – (40). We focus on the particular expression in equation (40), and others would be similar. Note first that

$$\frac{si}{n^2} - \frac{\bar{s}\bar{i}}{\bar{n}^2} = \frac{1}{n^2\bar{n}^2}\left[(i - \bar{i})s\bar{n}^2 + \bar{i}\bar{n}^2(s - \bar{s}) + \bar{i}\bar{s}(\bar{n} - n)(\bar{n} + n)\right]$$

In the following, for simplicity we write $Q = \frac{si}{n^2}$ with a similar definition for $\bar{Q}$, and we rely on bounds for the state variables established in Theorem 3.1.

$$
\begin{aligned}
&\left[\frac{si}{n^2}(\phi_2 - \phi_1) - \frac{\bar{s}\bar{i}}{\bar{n}^2}(\bar{\phi}_2 - \bar{\phi}_1)\right](\phi_3 - \bar{\phi}_3) \\
&= \left\{\phi_2\frac{1}{n^2\bar{n}^2}\left[(i - \bar{i})s\bar{n}^2 + \bar{i}\bar{n}^2(s - \bar{s}) + \bar{i}\bar{s}(\bar{n} - n)(\bar{n} + n)\right]\right. \\
&\quad -\bar{Q}(\phi_2 - \bar{\phi}_2) + \bar{Q}(\bar{\phi}_1 - \phi_1) \\
&\quad \left.-\phi_1\frac{1}{n^2\bar{n}^2}\left[(i - \bar{i})s\bar{n}^2 + \bar{i}\bar{n}^2(s - \bar{s}) + \bar{i}\bar{s}(\bar{n} - n)(\bar{n} + n)\right]\right\}(\phi_3 - \bar{\phi}_3) \\
&\leq C(\Psi + \Phi), \tag{44}
\end{aligned}
$$

where C depends on the bounds for n, and by extension s, i, and r, and the bounds for ϕ_1 and ϕ_2.

Equations (35) and (36) have terms containing $\frac{si}{n}$ and can be bounded similarly. Using the equality

$$\frac{si}{n} - \frac{\bar{s}\bar{i}}{\bar{n}} = \frac{s}{n}(i - \bar{i}) + \frac{s\bar{i}}{n\bar{n}}(\bar{n} - n) + (s - \bar{s})\cdot\frac{\bar{i}}{\bar{n}}$$

the following bound can be shown

$$[(i - \bar{i}) - (s - \bar{s})]\left[\frac{si}{n} - \frac{\bar{s}\bar{i}}{\bar{n}}\right] dt \leq C\left((i - \bar{i})^2 + (n - \bar{n})^2 + (s - \bar{s})^2\right), \tag{45}$$

with C dependent on bounds for s, i, and r.

A bound for $(\nu - \bar{\nu})^2$ requires

$$
\begin{aligned}
&\left[s\bar{r}^m(\phi_1 - \phi_3) - \bar{s}r^m(\bar{\phi}_1 - \bar{\phi}_3) \right]^2 \\
&= \left[s\bar{r}^m(\phi_1 - \bar{\phi}_1) + (s\bar{r}^m - \bar{s}\bar{r}^m + \bar{s}\bar{r}^m - \bar{s}r^m)\bar{\phi}_1 + s\bar{r}^m(-\phi_3 + \bar{\phi}_3) + \right. \\
&\leq 11 \left[s^2\bar{r}^{2m}(\phi_1 - \bar{\phi}_1)^2 + s^2\bar{r}^{2m}(\bar{\phi}_3 - \phi_3)^2 + \bar{r}^{2m}(s - \bar{s})^2(\bar{\phi}_1^2 + \bar{\phi}_3^2) \right. \\
&\qquad\qquad \left. + \bar{s}^2(\bar{r} - r)^2(\bar{\phi}_1^2 + \bar{\phi}_3^2)(m\max(r,\bar{r})^{m-1})^2 \right].
\end{aligned}
$$

The expresson $(\nu - \bar{\nu}^2)$ can be approximated following [9, 18] obtaining

$$
\begin{aligned}
(\nu - \bar{\nu})^2 &= \left(\frac{e^{2\alpha t}s(\phi_1 - \phi_3)}{\frac{2B_2}{K^m}e^{m\alpha t}r^m} - \frac{e^{2\alpha t}\bar{s}(\bar{\phi}_1 - \bar{\phi}_3)}{\frac{2B_2}{K^m}e^{m\alpha t}\bar{r}^m} \right)^2 \\
&\leq 11 \left[\frac{K^m e^{(2-m)\alpha t}e^{m\alpha t}}{2B_2 R^m} \right]^2 \cdot \\
&\qquad \left[s^2(\phi_1 - \bar{\phi}_1)^2 + s^2(\bar{\phi}_3 - \phi_3)^2 + (s - \bar{s})^2(\bar{\phi}_1^2 + \bar{\phi}_3^2) \right] \\
&\quad +11m^2 \left[\frac{K^m e^{(2-m)\alpha t}e^{m\alpha t}e^{\alpha t}}{2B_2 \min(R, \bar{R})^m \max(R, \bar{R})} \right]^2 \cdot \\
&\qquad \left[\bar{s}^2(\bar{r} - r)^2(\bar{\phi}_1^2 + \bar{\phi}_3^2) \right] \\
&\leq 11 \left[\frac{K^m e^{2\alpha t}}{2B_2} \right]^2 \cdot \\
&\qquad \left[\max(s)^2[(\phi_1 - \bar{\phi}_1)^2 + (\bar{\phi}_3 - \phi_3)^2] + (s - \bar{s})^2 4\max(\phi_1, \phi_3)^2 \right] \\
&\quad +11m^2 \left[\frac{K^m e^{3\alpha t}}{2B_2} \right]^2 \cdot \\
&\qquad \left[\max(s)^2 4\max(\phi_1, \phi_3)^2(\bar{r} - r)^2 \right] \\
&\leq C e^{6\alpha t} \left[(\phi_1 - \bar{\phi}_1)^2 + (\bar{\phi}_3 - \phi_3)^2 + (s - \bar{s})^2 + (\bar{r} - r)^2 \right] \qquad (46)
\end{aligned}
$$

with C dependent on bounds for s, ϕ_1 and ϕ_3 and that $R > 1$ which was assumed previously in the derivation of necessary conditions.

A bound on the second control depends on bounds for ϕ_3 and i:

$$
\begin{aligned}
(\tau - \bar{\tau})^2 &= \left(\frac{e^{2\alpha t}}{2B_3} \left[i(\phi_2 - \phi_3) - \bar{i}(\bar{\phi}_2 - \bar{\phi}_3) \right] \right)^2 \\
&\leq C e^{4\alpha t} \left[(i - \bar{i})^2 + (\phi_2 - \bar{\phi}_2)^2 + (\phi_3 - \bar{\phi}_3)^2 + (i - \bar{i})^2 \right] \qquad (47)
\end{aligned}
$$

The only form that remains to be considered is from equation (40):

$$
\begin{aligned}
&-\frac{mB_2}{K^m} \int_0^T e^{\alpha n t} \left[\nu^2 r^{m-1} - \bar{\nu}^2 \bar{r}^{m-1} \right] (\phi_3 - \bar{\phi}_3)\, dt \\
&\qquad \leq \frac{mB_2}{K^m} e^{m\alpha T} \int_0^T \left[\bar{\nu}^2(\bar{r} - r) \max(\bar{r}, r)^{m-2} \right. \\
&\qquad \leq \frac{mB_2}{K^m} e^{m\alpha T} \int_0^T \left. +(\bar{\nu}^2 - \nu^2)r^{m-1} \right] (\phi_3 - \bar{\phi}_3)\, dt \qquad (48)
\end{aligned}
$$

$$\leq \tilde{C}e^{2\alpha T} \int_0^T \left[\bar{\nu}^2(\bar{r} - r)\max(\bar{R}, R)^{m-2} \right.$$

$$\left. + (\bar{\nu}^2 - \nu^2)R^{m-1} \right] (\phi_3 - \bar{\phi}_3)\, dt$$

$$\leq Ce^{2\alpha T} \int_0^T \left[(\bar{r} - r)^2 + (\bar{\nu}^2 - \nu^2)^2 + (\phi_3 - \bar{\phi}_3)^2 \right]\, dt$$

where C depends on bounds for ν, R and the parameters m, B_2, K.

The bounds (41) – (48) are applied to the sum of the equations (35) – (40), with the latter three modified as described previously, deriving an inequality of the form:

$$\frac{1}{2}\left[\Psi(T) + \Phi(0)\right] + \alpha \int_0^T \left[\Psi(t) + \Phi(t)\right]\, dt$$

$$\leq \tilde{C} \int_0^T \left[\Psi(t) + \Phi(t)\right]\, dt + \hat{C}e^{6\alpha T} \int_0^T \left[\Psi(t) + \Phi(t)\right]\, dt$$

with the constants dependent on the parameter values and the established bounds for the state and adjoint variables. Rearranging this becomes

$$\frac{1}{2}\left[\Psi(T) + \Phi(0)\right] + (\alpha - \tilde{C} - \hat{C}e^{6\alpha T}) \int_0^T \left[\Psi(t) + \Phi(t)\right]\, dt \leq 0. \qquad (49)$$

We now choose α so that

$$\alpha > \tilde{C} + \hat{C}$$

and note that $\frac{\alpha - \tilde{C}}{\hat{C}} > 1$. Subsequently choose T so that

$$T < \frac{1}{6\alpha} \ln\left(\frac{\alpha - \tilde{C}}{\hat{C}}\right).$$

Then:

$$6\alpha T < \ln\left(\frac{\alpha - \tilde{C}}{\hat{C}}\right) \Rightarrow \exp(6\alpha T) < \frac{\alpha - \tilde{C}}{\hat{C}}.$$

It follows that $\alpha - \tilde{C} - \hat{C}\exp(6\alpha T) > 0$, so inequality (49) can hold if and only if for all $t \in [0, T]$

$$s(t) = \bar{s}(t), i(t) = \bar{i}(t), r(t) = \bar{r}(t),$$
$$\phi_1(t) = \bar{\phi}_1(t), \phi_2(t) = \bar{\phi}_2(t), \phi_3(t) = \bar{\phi}_3(t);$$

or,

$$S(t) = \bar{S}(t), I(t) = \bar{I}(t), R(t) = \bar{R}(t),$$
$$\lambda_1(t) = \bar{\lambda}_1(t), \lambda_2(t) = \bar{\lambda}_2(t), \lambda_3(t) = \bar{\lambda}_3(t);$$

establishing the uniqueness of the optimal control. $\qquad \square$

E-mail address: `hgaff@odu.edu`
E-mail address: `elsa.schaefer@marymount.edu`