Statistical Communications in Infectious Diseases

Manuscript 1020

Optimal Dynamic Policies for Influenza Management

Michael Ludkovski, *University of California, Santa Barbara*
Jarad Niemi, *University of California, Santa Barbara*

Optimal Dynamic Policies for Influenza Management

Michael Ludkovski and Jarad Niemi

Abstract

Management policies for influenza outbreaks balance the expected morbidity and mortality costs versus the cost of intervention policies. We present a methodology for dynamic determination of optimal policies in a completely observed stochastic compartmental model with parameter uncertainty. Our approach is simulation-based and searches the full set of sequential control strategies. For each time point, it generates a policy map describing the optimal intervention to implement as a function of outbreak state and Bayesian parameter posteriors. As a running example, we study a stochastic SIR model with isolation and vaccination as two possible interventions. Numerical simulations based on a classic influenza outbreak are used to explore the impact of various cost structures on management policies. Comparisons demonstrate the realized cost savings of choosing interventions based on the computed dynamic policy over simpler decision rules.

KEYWORDS: stochastic SIR model, sequential Bayesian inference, stochastic control, regression Monte Carlo

Author Notes: We are grateful to the anonymous referees and the associate editor for their careful reading of the manuscript that led to much improvement from an earlier version. We thank Bobby Gramacy for many useful discussions.

1 Purpose

Seasonal human influenza outbreaks affect the US annually and cost an estimated $90 billion per year (Meltzer et al., 1999). Flu surveillance by public health agencies, such as the Centers for Disease Control and Prevention and the World Health Organization, attempts to identify an impending outbreak and then counter-act with timely public advisories and interventions, such as vaccinations and quarantines. The objective of minimizing the outbreak's impact requires balancing the cost of preventive measures versus the cost of disease morbidity and mortality. Decision making is required under significant uncertainty since the precise outbreak dynamics, transmission rates, and effects of interventions are all unknown.

The research direction we present aims to make optimal sequential decisions in the face of such outbreaks. A disease propagation model is assumed with complete observations but unknown underlying parameters. As data are collected, mitigation policies can be implemented which effectively modify the outbreak parameters. Decisions are made adaptively and dynamically based on the costs involved in policy choices and outbreak severity. We consider an analysis which is appropriate for a small closed population, such as a boarding school (Murray, 2002, ch. 10), a fixed rural locality, or a college campus.

Statistical control of epidemics focuses on two major issues: 1) sequential estimation of unknown outbreak parameters and 2) designing control strategies that incorporate the uncertainty in parameter estimation. An extensive body of literature (Cintron-Arias et al., 2009, and references therein) treats parameter estimation of outbreaks, in particular estimating the basic reproductive ratio of the disease. In many cases, the nonlinear dynamics of the outbreak require computationally intensive estimation methods, such as Markov chain Monte Carlo (O'Neill, 2002, Elderd et al., 2006, Martínez-Beneito et al., 2008) and sequential Monte Carlo (Dukic et al., 2010). In larger populations, outbreak spread is described through hierarchical models that subdivide the population into households and cities (Ball and Neal, 2002), introducing heterogenous mixing and making estimation even more challenging. Recently, real-time flu tracking software, such as Google Flu Trends, has been used to obtain sequential outbreak estimates by Dukic et al. (2010) and Niemi (2009).

The second issue consists in designing control strategies to mitigate the outbreak impact. While a comprehensive analysis is available for deterministic models (see e.g. the reviews in Hethcote (2000) and Tennenbaum (2008)), it is only recently that the stochastic nature of the outbreak and parameter uncertainty have been simultaneously considered. Ball and Lyne (2002), as well

as Patel et al. (2005), study optimal allocation of limited vaccine quantities among households. Clancy and Green (2007) consider the impact of parameter uncertainty on optimal vaccination strategies in a static setting where parameter posteriors are not updated. Alternatively, Merl et al. (2009) search for optimal vaccination strategies in a fully Bayesian setup but make decisions only at one time-point rather than as data are collected. More formal stochastic control formulations have been considered by Cai and Luo (1994) and Tanner et al. (2008) who use the framework of stochastic programming and Markov decision processes. From a public health perspective, many studies have described realistic guides for containment strategies of specific outbreaks, such as the 2009 H1N1 pandemic (Halloran et al., 2008, Germann et al., 2006, Elderd et al., 2006, Yang et al., 2009).

In this paper we propose a new methodology to address the problem of sequential management of flu outbreaks. We do not restrict *a priori* the set of possible control strategies. Instead our algorithm finds the approximately optimal decision rule at each decision time based on the current outbreak state and parameter posteriors. We continue this manuscript by introducing the stochastic SIR model and sequential Bayesian inference in Section 2. Section 3 introduces the cost structure and the dynamic control methodology which utilizes dynamic programming and regression Monte Carlo. Section 4 implements the methodology on a classic influenza outbreak example under various cost structures. This section also compares our dynamic policy to simpler decision rules. Finally, we conclude with some discussion and future work in Section 5.

2 Disease Propagation Model

For the disease propagation model, we consider the general stochastic SIR outbreak model (Andersson and Britton, 2000, ch. 2). We assume a single-level compartmental model that has three mutually exclusive categories of individuals: susceptible (S), infected (I), and removed (R). We consider a constant population size $N = S_t + I_t + R_t$ with no immigration or deaths. Generally, after getting infected, individuals acquire immunity against the particular flu strain for a given year, so that we may assume removed individuals are immune. Accordingly, an outbreak is described through three possible transitions:

$S \to I$: a susceptible individual becomes sick,
$I \to R$: an infected individual recovers and has immunity from the disease, and
$S \to R$: a susceptible individual is vaccinated and has immunity from the disease.

The rate of transitions depend on unknown fixed parameters, as well as the current number of individuals in each category. The system evolves according to a Markovian structure on the outbreak progress which implies that the triple (S_t, I_t, R_t) is a continuous-time Markov chain on the domain $\{(s,i,r) \in \mathbb{Z}_+^3 : s+i+r = N\}$. This Markov process is governed by the transition equations:

$$\begin{aligned}
\mathbb{P}[(S_{t+\delta}, I_{t+\delta}, R_{t+\delta}) = (s-1, i+1, r) \big| (S_t, I_t, R_t) = (s,i,r)] = \quad & k_I^{\phi_t^I} \beta i s/N \, \delta + o(\delta), \\
\mathbb{P}[(S_{t+\delta}, I_{t+\delta}, R_{t+\delta}) = (s, i-1, r+1) \big| (S_t, I_t, R_t) = (s,i,r)] = \quad & \gamma i \, \delta + o(\delta), \\
\mathbb{P}[(S_{t+\delta}, I_{t+\delta}, R_{t+\delta}) = (s-1, i, r+1) \big| (S_t, I_t, R_t) = (s,i,r)] = \quad & \phi_t^V k_V s \, \delta + o(\delta),
\end{aligned} \tag{1}$$

with all other transitions having probabilities of $o(\delta)$. The rate of a susceptible individual becoming infected is proportional to the product $i \cdot s$ representing the possible pairings between infected and susceptible individuals. Each infected individual stays sick for an exponentially distributed amount of time (independent of everything else) with rate γ, so that the expected duration of individual infection is $1/\gamma$ time units. Finally, susceptible individuals get vaccinated at rate k_V if vaccination is currently implemented. Equations (1) are rigorously interpreted in the limit $\delta \to 0$ and the small-o notation means that those terms become negligible for δ small enough.

The constants β and γ are intrinsic disease parameters that describe respectively the infectiousness and duration of infection. The quantities $\phi^I, \phi^V \in \{0,1\}$ are *controls* that can be used respectively to lower infectiousness through *isolation* and provide immunity through *vaccination*. Initially the controls are zero, but can be switched on which either lowers the infection rate ($0 < k_I < 1$ known) if $\phi^I = 1$ or makes vaccine available ($k_V > 0$ known) if $\phi^V = 1$. More discussion of the ϕ-controls is in Section 3.

The outbreak ends stochastically when there are no longer any infecteds left. We denote the outbreak end-time by $\tau := \inf\{t \geq 0 : I_t = 0\}$. Since N is finite and the transition rates are all bounded away from zero, the outbreak will end with probability 1, $\mathbb{P}(\tau < \infty) = 1$.

Consider briefly the situation where vaccination is unavailable, $\phi^V = 0$. Then it is well-known that a key parameter for the outbreak is the basic reproductive number $\mathcal{R}_0 := \beta/\gamma$. If $\mathcal{R}_0 > 1$ then the number of new infecteds will on average initially outpace the number of removed individuals and an outbreak occurs. Conversely, if $\mathcal{R}_0 < 1$, the probability of a major outbreak is negligible. The basic reproductive number determines the probability of an outbreak, which is approximately $1 - 1/\mathcal{R}_0$ for $\mathcal{R}_0 > 1$, and also controls the expected number of infecteds over the course of the outbreak, $\mathbb{E}[R_\tau]$, which can

be computed approximately from the Kermack-McKendrick equation (Bailey, 1975):

$$\log((N - \mathbb{E}[R_\tau])/S_0) + \mathcal{R}_0 \mathbb{E}[R_\tau]/N = 0. \tag{2}$$

It is important to remember that due to stochastic fluctuations the distribution of total infecteds R_τ is bimodal for all $\mathcal{R}_0$: one mode if the outbreak randomly dies out quickly and another if the outbreak is sustained.

2.1 Parameter estimation

The outbreak spread is crucially determined by the outbreak parameters β and γ which determine the basic reproductive ratio $\mathcal{R}_0$. These parameters describe the epidemiological properties of the flu strain and are generally unknown. Indeed, their estimation is one of the key statistical problems in epidemiology. While methods exist for estimating these parameters once the outbreak is over (see for instance (Andersson and Britton, 2000, ch. 9)), for the purposes of outbreak control, an online estimation procedure is advantageous. Indeed, as outbreak data are collected, officials in charge will be computing latest estimates of β, γ, and $\mathcal{R}_0$, as well as their uncertainty, to enact policy decisions.

Such considerations suggest a Bayesian approach whereupon full posterior distributions of the outbreak parameters are sequentially computed. In the model of equation (1), we assume that priors for β and γ are given at $t=0$ and are updated sequentially based on continuous observation of the outbreak state vector. Since the outbreak state is discrete and changes only occur at infection/removal/vaccination times, this is equivalent to assuming that the officials observe all transitions and transition times. Because $R_t = N - S_t - I_t$, the number of removed individuals can always be inferred from the (S_t, I_t)-pair and will not be mentioned again.

Independent gamma priors are assumed for β and γ, i.e. $p(\beta, \gamma) = Ga(\beta; a_\beta, b_\beta)\, Ga(\gamma; a_\gamma, b_\gamma)$, where $Ga(x; a, b)$ denotes that x is a gamma-distributed random variable with mean a/b and variance a/b^2. When transitions and their times are known, these priors are conjugate and the posteriors can be computed at any time t using the following formula:

$$\begin{aligned} p\left(\beta, \gamma \,\middle|\, S_{(0,t)}, I_{(0,t)}, \phi^I_{(0,t)}\right) &= Ga\left(\beta; a_\beta + r_1, b_\beta + \frac{1}{N}\int_0^t k_I^{\phi_t^I} I_t S_t dt\right) \\ &\quad \times Ga\left(\gamma; a_\gamma + r_2, b_\gamma + \int_0^t I_t dt\right) \end{aligned} \tag{3}$$

where r_1 is the number of $S \to I$ transitions and r_2 is the number of $I \to R$ transitions (Wilkinson, 2006, ch. 10). Clearly the posteriors depend on

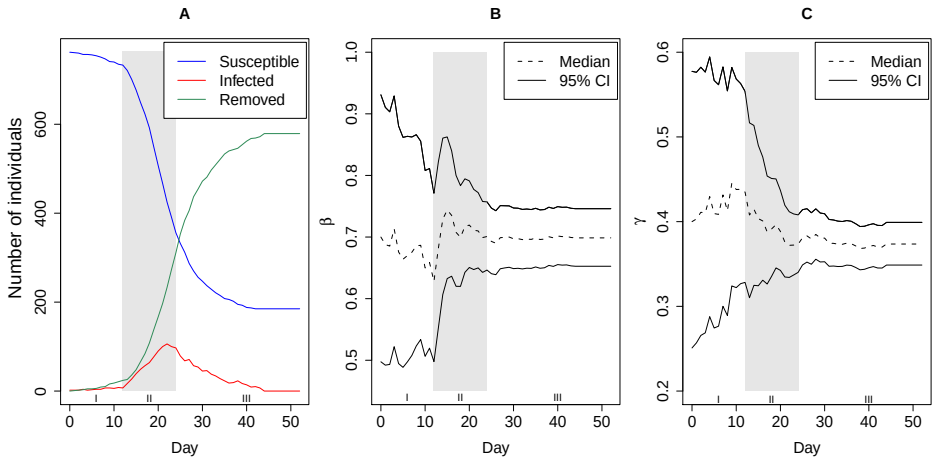

Figure 1: (A) A realization of the model in equation (1) with $\beta = 0.7$ and $\gamma = 0.4$ with no controls, i.e. $\phi_t^I = 0$ and $\phi_t^V = 0 \,\forall\, t$, along with sequential point-wise medians and 95% credible intervals for sequential inference on β (B) and γ (C) assuming independent gamma priors $Ga(\beta; 28, 40)$ and $Ga(\gamma; 16, 40)$ for β and γ, respectively.

both the state of the outbreak and whether the isolation control is in effect. The integrals within these posteriors are actually summations since both the controls and the state of the system change by discrete quantities at distinct, observed times. The uncertainty of these parameters decreases as the integrals in equation (3) increase and stays constant once $I_t = 0$.

Figure 1 provides a simulation of the SIR model based on the method described in Appendix A, as well as inference on the unknown fixed parameters as described in this section. The outbreak progress can be divided into three phases. In phase I (see Figure 1), the outbreak is just beginning and the absolute number of infecteds grows slowly (since I_0 was small). At this point officials are waiting for information about the outbreak parameters before making major decisions and may undertake cheaper preventive measures. Also, in some scenarios the outbreak will cease on its own and no controls become necessary.

In phase II (gray area in Fig 1), the number of infecteds grows rapidly and parameter learning occurs. The figure illustrates that uncertainty in β decreases quicker than that in γ since the $S \to I$ transition necessarily occurs before the $I \to R$ transition. Based on the number of infecteds and parameter posteriors (especially that of $\mathcal{R}_0$), officials choose interventions based on their public health policy. The timing of phase II will vary across scenarios, since

often the outbreak will stay small for several periods before finally breaking out. The adaptive timing of policy decisions will also vary across scenarios and is central to optimizing expected costs. Furthermore, policies may be dynamically adjusted in the face of further information or unusual outbreak behavior.

In phase III, the outbreak dies out, parameter learning stops, and no new interventions are undertaken. The final posteriors for the two parameters in Figure 1 are $Ga(\beta; 605, 866)Ga(\gamma; 595, 1293)$ indicating the lack of information provided by the prior relative to the data for this outbreak that ultimately led to 579 (76%) of the individuals being infected.

3 Dynamic Policy Maps

The public health objective is to minimize the total cost of the outbreak. The spread of the outbreak can be mitigated by undertaking *interventions*, however these interventions are costly. Thus, a trade-off exists between the cost of morbidity and mortality and the cost of public action. If the outbreak is mild, the economically efficient course of action is to do nothing. In contrast, if the outbreak is severe, it may be most efficient to act as soon as possible to limit total number of infecteds and the associated costs. A major tension exists between these actions under uncertainty, when the outbreak parameters are unknown and estimated dynamically. The "wait-and-see" option becomes valuable allowing for parameter learning to take place before committing to any particular course of action.

3.1 Costs

We capture the basic costs of the outbreak through the rate $c_1(I_t)$, i.e. the instantaneous cost for having I_t infecteds. In the base case, we take $c_1(\cdot)$ to be linear; however $c_1(\cdot)$ should generally be convex to account for additional costs of severe epidemics.

As mentioned above, costs are incurred for any intervention that interferes with a disease's natural course. The intervention cost consists of a fixed cost (significant in practice) that is paid once when the action is started, and a variable cost that may depend on the latest number of actual infecteds/susceptibles that are treated and which is expressed as a rate to be paid continuously. To fix ideas, we focus on two major policies: isolation and vaccination. When activated, the isolation control $\phi_t^I \in \{0, 1\}$ reduces the infectiousness rate between susceptible and infected individuals. The cost of

isolation is taken to be linear in the current number of infecteds, i.e. $(c_2 I_t + c_3)$, plus a one-time upfront cost, $c_6(t)$. When activated, the vaccination control $\phi_t^V \in \{0,1\}$ shrinks the susceptible pool by having direct $S \to R$ transitions. Note that vaccination is not instantaneous but is described again through a transition rate. This captures the phenomenon of vaccinations administered over a period of time, with varying efficacy, population response, and time to immunization. Equation (1) implies that vaccination is available to the entire susceptible population with a constant vaccination rate for each susceptible individual (thus total vaccination rate is proportional to S_t). Similarly, we assume the cost of vaccination is linear in the number of remaining susceptibles, i.e. $(c_4 S_t + c_5)$. Moreover, since quick development and manufacturing of vaccines is costly, introduction of a vaccination policy has a time-dependent fixed cost $c_7(t)$ which is decreasing in t.

The overall control at any time t is a pair $\phi_t := (\phi_t^I, \phi_t^V) \in \{0,1\}^2$ and ϕ is the set of controls over all times. Let T be the endpoint under consideration, then the total cost for an outbreak on the horizon $[s,T]$ is

$$C_s^T(\phi) = \int_s^T \left[c_1(I_t) + (c_2 I_t + c_3)\phi_t^I + (c_4 S_t + c_5)\phi_t^V \right] dt \\ + c_6(t_I)\mathrm{I}(\phi_s^I = 0) + c_7(t_V)\mathrm{I}(\phi_s^V = 0) \tag{4}$$

where $t_I = \inf\{t \in [0,T] : \phi_t^I = 1\}$ and $t_V = \inf\{t \in [0,T] : \phi_t^V = 1\}$ indicate when isolation and vaccination are initiated and $c_6(\infty) = c_7(\infty) = 0$ if the control is never initiated. The last two terms correspond to fixed costs incurred when starting isolation or vaccination, respectively. We take $c_6(t_I)$ to be constant and $c_7(t_V)$ to be decreasing (e.g. exponentially) in t_V. For simplicity, we limit the possible controls to have vaccination and isolation turned on at most once.

3.2 Optimal policy choices

Even for constant controls ϕ^I, ϕ^V and known outbreak parameters β, γ, the nonlinear dynamics in equation (1) imply that expected costs over the outbreak lifetime, $\mathbb{E}\left[C_0^T(\phi) | S_0 = s_0, I_0 = i_0\right]$, cannot be computed in closed-form. In the case without vaccination, good approximations are generally available relying on the Kermack-McKendrick formula in equation (2) and its extensions, see Bailey (1975); however even those are complicated to use under parameter uncertainty due to additional integrals over the parameter distributions. With vaccination and dynamic controls, there is no hope to compute equation (4) analytically.

In order to compute an (approximately) optimal disease intervention policy, we use a stochastic control framework. The time-t parameter posteriors in equation (3) can be summarized by the updated shape and rate hyperparameters for the associated gamma distributions. Let Z_t represent this quadruple which is a continuous-time stochastic process taking values in the space $\mathbb{Z}_+^2 \times \mathbb{R}_+^2$ and, under the assumptions of our model, is Markov. The overall stochastic outbreak is summarized by the Markov sextuple $X_t := (S_t, I_t, Z_t)$.

The evolution of X_t is affected by the policy vector ϕ which modifies both the dynamics of (S_t, I_t) and the inference of β as described in Section 2.1. The policy vector also directly impacts the cost functional $C_s^T(\phi)$ and the direct cost of outbreak $\int_0^T c_1(I_t)dt$. The policy objective can be stated as

$$\inf_{\phi_{(0,T)}} \mathbb{E}\left[C_0^T(\phi)\middle|\ X_0 = x_0\right], \qquad \phi_0 = (0,0), \tag{5}$$

where the sub- and super-scripts on C are used to indicate the expectation is taken over possible future outbreak realizations from time 0 to T, i.e. $S_{(0,T]}, I_{(0,T]}$. We assume control decisions are undertaken at specified intervals although other frequencies can be easily accommodated (in principle we can even consider continuous-time controls). Thus our control space becomes $\phi_{0:T-1} = (\phi_0, \phi_1, \ldots, \phi_{T-1})$ where ϕ_t gets implemented at time t and affects the dynamics from t to $t+1$. Since the total control space is finite (discrete in time and discrete at each time-point), the infimum in equation (5) is achieved and an optimal management policy exists.

Despite the low number of possible controls $\phi_t \in \{0,1\}^2$ at each time point, the full size of this set over all times (4^T) is too large for direct enumeration. In addition, a fixed policy vector $\phi_{0:T-1}$ that is optimal for one outbreak is unlikely to also be optimal for a new outbreak. Therefore, the objective is to dynamically set $\phi_{0:T-1}$ to minimize the expected total cost $C_0^T(\phi)$ defined in equation (4). The control ϕ_t is random and will be determined adaptively based on time-t information X_t and the currently implemented policy ϕ_{t-1}. This function, $\Phi(t, x_t, \phi_{t-1}) \mapsto \phi_t$, is termed a *policy map*.

In the next two sections we describe how to create this policy map in two general steps. The first step uses dynamic programming to create the maps in reverse chronological order. The second step uses regression Monte Carlo to create the map at a time t if all maps from time $t+1$ to T have already been created.

3.2.1 Dynamic programming

Due to the Markov nature of the model and inference, it is easier (and computationally more tractable) to visualize the global optimization problem above as a sequential one-step optimization. Consider a family of optimizations indexed by their starting time t, policy regime ϕ_{t-1}, and state vector x_t. Define

$$\mathfrak{C}_t(\phi_t) = \mathfrak{C}(t, x_t, \phi_t) = \min_{\phi_{t+1:T}} \mathbb{E}\left[C_t^T(\phi)\middle| x_t\right]. \tag{6}$$

We call $\mathfrak{C}_t$ the *continuation cost* at time t; its interpretation is that of minimal expected future costs assuming that an admissible optimal dynamic policy is followed after t and the (possibly sub-optimal) policy at time t is ϕ_t. The original objective is computing $\mathfrak{C}_0 = \mathfrak{C}(0, x_0, \phi_0)$. The deterministic terminal condition is $\mathfrak{C}_T = 0$ by definition.

Denote by $F(\phi)$ the set of possible policies given that previous policy was ϕ. At any time t, this set could include all possible policies or could be restricted to some subset. Due to the Markovian nature of the state vector, we may apply Bellman optimality principle (Bertsekas, 2005, Fleming and Soner, 1993) to conclude that

$$\mathfrak{C}_t(\phi_t) = \mathbb{E}\left[C_t^{t+1}(\phi_t)\right] + \min_{\phi_{t+1}\in F(\phi_t)} \mathbb{E}\left[\mathfrak{C}_{t+1}(\phi_{t+1})\right], \tag{7}$$

where the second expectation is taken over just the outbreak states at $t+1$, i.e. $p(X_{t+1}|x_t, \phi_t)$. Equation (7) expresses the intuitive idea that if we know the best adaptive policy on $[t+1, T]$ we can extend it to the best policy on $[t, T]$ by optimizing the local policy at the next time-step.

This optimization is global (rather than myopic) because $\mathfrak{C}_{t+1}(\phi_{t+1})$ not only summarizes the entire future policy but also takes into account the effect of today's policy on future system states $S_{(t+1:T]}, I_{(t+1:T]}$. Moreover, implementing (7) is practical since $F(\phi_t)$ is small, and therefore one can directly enumerate over all possible choices. Once $\mathfrak{C}_t(\phi_t)$ is known for all ϕ_t, we can obtain the policy map $\Phi(t, x_t, \phi_{t-1})$ by taking the ϕ_t that minimizes $\mathfrak{C}_t(\phi_t)$ at the particular values of x_t, ϕ_{t-1}.

The overall dynamic programming procedure consists of iteratively solving equation (7) backward in time starting from the known terminal condition $\mathfrak{C}_T = 0$. While creation of these policy maps is computationally demanding, they are created only once prior to the outbreak. Once completed, the maps are used directly in policy-making for any future realized outbreak scenario. Namely, one sequentially collects outbreak data and, at each time t, selects the intervention given by $\Phi(t, x_t, \phi_{t-1})$.

3.2.2 Regression Monte Carlo

In general, a variety of methods are available for evaluating (7), see Ludkovski (2005) for a discussion. We stress that no exact solution of (7) is possible and all methods are approximate since no analytic formulas are available for expected outbreak costs. Even if some policies admit analytic expressions for their expected costs, it is impossible to *a priori* prove that they constitute the optimal strategy. Therefore, to be robust, the decision-maker should be agnostic about what types of rules *may* be optimal and search a large enough family of potential policies first.

To resolve these issues, we apply a simulation-based algorithm called regression Monte Carlo (RMC), introduced to stochastic control by Longstaff and Schwartz (2001) and more formally in Egloff (2005), and studied in the context of partially observed models by Ludkovski (2009). The key idea of RMC is to apply the backward recursion of equation (6) on a stochastic mesh corresponding to a collection of simulated paths $x_{0:T}^{(k)} \in X_{0:T}$ for $k = 1, \ldots, K$. Recall that in a basic Monte Carlo method, the one-step-ahead expectations in equation (7) can be computed by generating n scenarios starting with the epidemic state (ϕ_t, x_t) at date t and then taking the empirical average,

$$\mathbb{E}\left[\mathfrak{C}_{t+1}(\phi_{t+1}) \,\middle|\, x_t^{(k)}\right] \approx \frac{1}{n}\sum_{j=1}^{n} \tilde{c}_{t+1}^{[j]},$$

where $\tilde{c}_{t+1}^{[j]}$, $j = 1, \cdots, n$ are realizations of $\mathfrak{C}_{t+1}(\phi_{t+1})|x_t^{(k)}$. Such a scheme is computationally very expensive however, as it requires a forest of Monte Carlo simulations for each point in the stochastic mesh. To reduce such computational costs, we generate just a single realization (or pathwise continuation cost) $\mathfrak{c}_{t+1}^{(k)}$ of $\mathfrak{C}_{t+1}(\phi_{t+1})|x_t^{(k)}$ for each point $x_t^{(k)}$ in the mesh. We then use cross-sectional information from other points $x_t^{(k)}$ in the mesh by regressing the $\mathfrak{c}_{t+1}^{(k)}$ against chosen outbreak statistics. In other words, we replace the above empirical averaging with a regression-based prediction for the right-hand-side of (7). We use ordinary least-squares regression due to the characterization of conditional expectation appearing in (7) as L^2-minimizer; practically other regressions can be applied to reduce variance.

The formula (7) is recursive, so that when applying it at date t, we use the (already computed) optimal intervention strategies on $[t+1, T]$ to obtain $\mathfrak{c}_{t+1}^{(k)}$. In addition, to avoid the rapid accumulation of errors due to repeated approximations in (7), a further improvement is to use this approximation only for the policy maps, rather than directly for continuation values.

Let us detail more formally the RMC procedure shown in Algorithm 1 in Appendix B.1. We begin by simulating a large outbreak scenario database $(x^{(k)}_{0:T})$ starting with known initial condition X_0 and a default policy map Φ^0 via repeated uses of Algorithm 2. In principle this default map is arbitrary, but choices closer to the actual optimal policy will have less extrapolation and therefore improved accuracy for a fixed number of simulations. This database represents a stochastic mesh upon which we create policy maps by recursively implementing equation (7) backward from $t = T$ to $t = 0$.

Assuming we have an optimal policy map from time $t+1$ to T, we now detail the construction of a policy map at time t. For all possible policies ϕ and each $x^{(k)}_t$, we use Algorithm 2 to re-simulate one forward outbreak realization $\tilde{x}^{(k,\phi)}_{t+1:T}$ starting from $x^{(k)}_t$, using control ϕ at t, and implementing future controls based on the already constructed policy maps from time $t+1$ to T. This yields a simulated continuation cost (for the period $[t, T]$), $\mathfrak{c}^{(k)}_t \sim \mathfrak{C}(t, x^{(k)}_t, \phi)$, for each starting point $(\phi, x^{(k)}_t)$. Note that in contrast to the original database $x^{(k)}_{t+1:T}$ that used Φ^0, this re-simulation $\tilde{x}^{(k)}_{t+1:T}$ uses the optimal map $\Phi(s, \cdot)$, $s \in [t + 1, T]$.

We choose relevant statistics of X_t to use as explanatory covariates, i.e. outbreak statistics, for the decision rules. For each ϕ, we regress the simulated continuation costs onto these statistics. Typical statistics we use are the current

- levels of S_t and I_t,
- posterior means and variances of β, γ, and $\mathcal{R}_0$,
- posterior means of rate transition functions, e.g. $\beta I_t S_t$, and
- posterior means of cost structure functions, e.g. I_t/γ.

Since the posteriors of outbreak parameters are independently gamma-distributed, all moments of β, γ and their ratios have closed-form expressions in terms of the respective shape and rate parameters.

Let $F(\phi_{t-1})$ contain the set of allowable policies if ϕ_{t-1} is implemented at time $t - 1$. If no restrictions are in place, $F(\cdot) = \{0, 1\}^2$. The policy map $\Phi(t, x_t, \phi_{t-1})$ is constructed for *all* x_t, ϕ_{t-1} by choosing the policy $\phi_t \in F(\phi_{t-1})$ with lowest regression predicted costs.

As with any regression, accuracy of predictions is highest in neighborhoods where many scenarios reside. But, to have a good approximation to the solution, we need the regression to be accurate in the neighborhood where the optimal X_t is likely to be found (e.g. if optimally managed outbreaks always stay below 50 infecteds, our forward-simulated scenarios should reflect this fact) — which is *a priori* unknown. In general, to improve accuracy, we could

increase the number of statistics used as regression bases; however to avoid over-fitting, the number of simulations would need to increase as well. Let us also recall that since the policy maps Φ are generated through regressions on a set of simulation paths, different MC runs will generate slightly different policy rules and slightly different expected costs. In the example below, we find that this standard error is less than 0.5%.

The RMC scheme above has several attractive features. First, it allows to fully capture the nonlinear system dynamics without any *a priori* discretization and utilizes statistics that are important for the optimization stage. Thus, any additional knowledge about the optimal control structure can be directly input into the regression step leading to improved accuracy. Second, it requires minimal design changes with respect to chosen outbreak dynamics, cost structures, and possible policy choices/sequences. Third, by re-using the information from multiple outbreak simulations, it increases computational efficiency. Finally, it allows transparent analysis for the incurred approximation errors and vast potential for variance reduction.

4 Boarding School Example

As an illustration of the methodology described above, we consider the case of a flu outbreak in an English boarding school described in Murray (Murray, 2002, ch. 10) and recently used by Merl et al. (2009). The population consisted of $N = 763$ students. As our initial state, we take $S_0 = 761, I_0 = 2, R_0 = 0$ so that the model begins with two infecteds and everyone else susceptible. For priors, we assume $p(\beta, \gamma) = Ga(\beta; 28, 40)Ga(\gamma; 16, 40)$ so that the expectations comply with the estimates Murray found and the variances result in reasonable outbreak profiles. With this prior, the resulting basic reproductive ratio of the outbreak $\mathcal{R}_0$ has a prior 95% probability interval of $[0.96, 3.35]$ which ranges from no outbreak $\mathcal{R}_0 < 1$ to severe outbreak with $\mathcal{R}_0 > 3$ that unchecked would infect over 80% of the population. The data in Murray (2002) exhibit a relatively short-lived flu outbreak over 25 days; accordingly in this example our time unit will be 1 day.

We normalize our costs to $c_1(I_t) = I_t$, so that each infected student costs 1 per infected-day and all other costs should be interpreted relative to this one. The other costs are associated with the three possible policies: wait-and-see, isolation, and vaccination. Under wait-and-see no policy actions are taken and it is the optimal strategy if the outbreak is non-existent or mild. Isolation restricts interaction among susceptibles and infecteds, e.g. canceling classes or quarantine. As a result, isolation lowers the infectiousness rate by a factor

of 2 ($k_I = 0.5$) and is an effective policy if $\mathcal{R}_0/2 = \beta/2\gamma < 1$. In addition, isolation costs have a relatively modest fixed upfront cost $c_6(t_I) = 400$ and a running cost that is linear in the current number of infecteds $c_2 I_t + c_3 = 7I_t + 20$. Vaccination moves susceptibles to the removed population at an average rate of 20% per day ($k_V = 0.2$), so that about half of all susceptibles are vaccinated within 3 days. Its administration cost is a constant $c_4 S_t + c_5 = 20$. In addition, vaccination is very costly to obtain initially and therefore carries a major *upfront* cost that is exponentially decreasing in time to account for vaccine transportation costs, i.e. a major sum can be paid to get the vaccine immediately. We take $c_7(t_V) = 2500e^{-(t_V-1)/7}$ so that the cost of obtaining a vaccine is 2500 at $t = 1$, about 1000 at $t = 7$ and less than 150 after three weeks $t = 21$. Because vaccination repeatedly lowers the fraction of susceptibles, it will defeat any outbreak, no matter how severe; however its upfront cost would be too high if the outbreak is not severe.

4.1 Boarding school optimal policy map

Typical uncontrolled outbreaks under the priors above start slowly in week 1 with about 25% of the cases never reaching critical mass for an epidemic curve. Epidemic outbreaks reach their pinnacles in weeks 2-3 and then slowly disappear in weeks 4-6. In over 95% of the cases, outbreaks are extinct after $T = 45$ days and we use that as our control horizon. From a control perspective, the crucial days are $4 \le t \le 16$ when an appropriate policy must be chosen to combat the outbreak. Since vaccination carries only upfront costs and always succeeds, we assume that any vaccination decision is permanent and additional isolation is never necessary. In contrast, if the isolation policy is not enough, it may be augmented or replaced with vaccination to stop further spread. To summarize, possible control transitions are $F((0,0)) = \{0,1\}^2$; $F((1,0)) = \{(1,0),(0,1),(1,1)\}$ and $F((0,1)) = (0,1)$, and $F((1,1)) = (1,1)$. To create an optimal policy map for this example, we ran the entire procedure outlined in Section 3.2 using $K = 40,000$ scenarios using 10 outbreak statistics at each time step: $\{I_t, I_t^2, S_t, \mathbb{E}_t[\mathcal{R}_0], \mathbb{E}_t[\mathcal{R}_0]^2, \mathbb{E}_t[\beta], \mathbb{E}_t[\gamma], \mathbb{E}_t[\beta] I_t S_t, I_t \mathbb{E}_t[1/\gamma], \mathbb{E}_t[\mathcal{R}_0] I_t\}$, where $\mathbb{E}_t$ denotes the time-t posterior mean. Algorithm 2 is written in C while Algorithm 1 is written in R (R Development Core Team, 2009). The resulting optimization takes about 2 hours on a 3GHz server and requires 500Mb of memory to run.

The full policy map is a function of all six dimensions of our model, as well as time, and therefore its visualization is difficult. A simplified version is shown in Figure 2 which shows a slice of this map for various number of

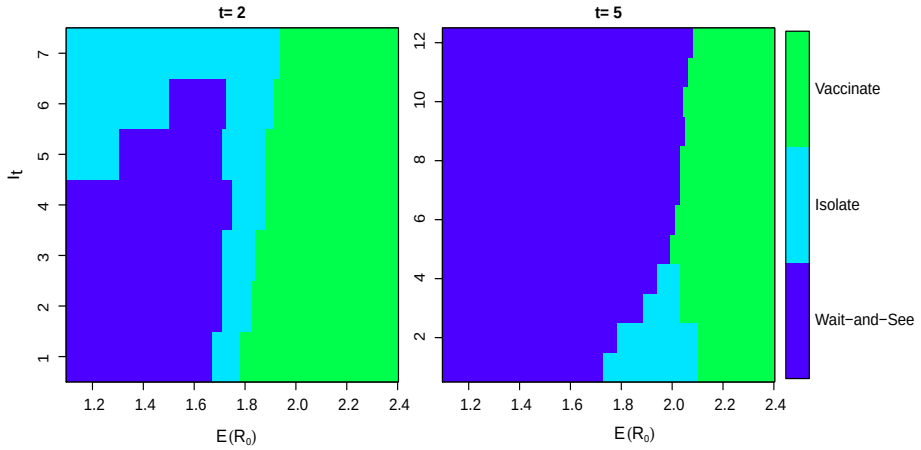

Figure 2: Optimal policy maps (left: $t=2$, right: $t=5$) as a function of current infecteds (discrete) and posterior mean of the basic reproductive number (continuous) assuming wait-and-see was previously implemented. For the outbreak states we take $p(\gamma, \beta|X_t) = Ga(\gamma; 40, 100)Ga(\beta; 40\mathbb{E}[\mathcal{R}_0|X_t], 100)$ and $S_t = 761 - \lceil 1.5I_t \rceil$ (left), $S_t = 759 - \lceil 1.5I_t \rceil$ (right panel). Note the different y-axis scales necessary due to the non-stationary behavior of the outbreak.

infecteds I_t and $\mathbb{E}_t[\mathcal{R}_0]$ with other dimensions at reasonable values. This map is included as an illustration, but its interpretation requires accounting for the relative probabilities of the locations on the map. For example, if I_t is small at time $t = 2$ then it is virtually impossible to have a large $\mathbb{E}_t[\mathcal{R}_0]$ since this requires long morbidity duration or high infectivity both of which would lead to large number of infecteds.

At $t = 2$, the outbreak is only starting and so the variability in I_2 and $\mathbb{E}_2[\mathcal{R}_0]$ is small; the optimal strategy is to wait-and-see if the outbreak appears to be moderate and $I_t \leq 4$, and to apply isolation if either $\mathbb{E}_2[\mathcal{R}_0]$ or I_t is moderate. Vaccination is still too costly and is only applied if the outbreak is severe, i.e. large $\mathbb{E}_2[\mathcal{R}_0]$. At $t=5$, cost of vaccination drops and it is applied whenever the posterior mean of $\mathcal{R}_0$ is large enough that isolation alone is insufficient. Isolation is now applied only if I_t is small and $\mathbb{E}_5[\mathcal{R}_0] \in [1.7, 2.1]$. While this region is small on the map, it is actually a relatively common outcome.

To demonstrate the use of this map, Figure 3 shows *sample paths* of four controlled outbreaks and their associated policy decisions. All four outbreaks had parameter priors as indicated above and the same true β, γ. Thus,

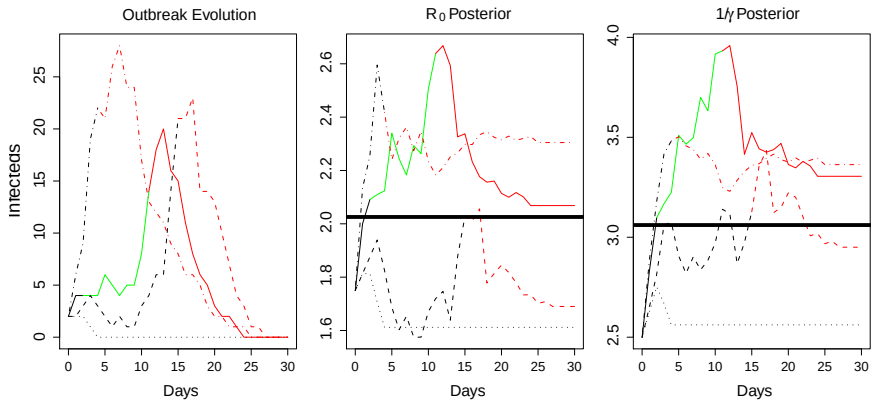

Figure 3: Four controlled outbreak simulations (solid, dotted, dashed, and dot-dash) with paths colored according to the current intervention: wait-and-see (black), isolation (green), and vaccination (red). The middle and right plots show posterior means for reproductive ratio ($\mathcal{R}_0$) and expected morbidity (γ^{-1}) for the same simulations with true parameter values (solid black).

all variability is entirely due to the stochastic SIR model fluctuations. The dotted simulation outbreak dies out on its own and therefore no intervention was chosen. The dashed outbreak initially has low posterior mean for $\mathcal{R}_0$ and therefore the wait-and-see policy is enacted until the number of infecteds and posterior mean for $\mathcal{R}_0$ rise. At $t = 15$, vaccination is cheap enough that it is immediately implemented to control the outbreak. The solid outbreak initially has only a few infecteds and the posterior means suggest a moderate outbreak impending. Isolation is enacted to keep the outbreak modest until vaccination is cost effective at $t = 12$. The dot-dash outbreak is a quickly rising epidemic and therefore vaccination is enacted at time 4 to control this outbreak. Isolation is never implemented in this scenario because it would have added an unnecessary cost.

To complement Figure 3, Figure 4 gives a high-level summary of the dynamic control strategy in this example based on $25,000$ realizations (each with parameters drawn from the prior). The left panel of Figure 4 shows the median and 95% quantile of the number of infecteds over time in the controlled (green) and uncontrolled (red) scenarios. During the first five days, control has a minor effect since policy actions are typically not yet taken. Overall, however, the optimal control strategy lowers the 95%-quantile of the peak number of infecteds from over 200 to under 70 and also dramatically reduces the median number of infecteds at all times.

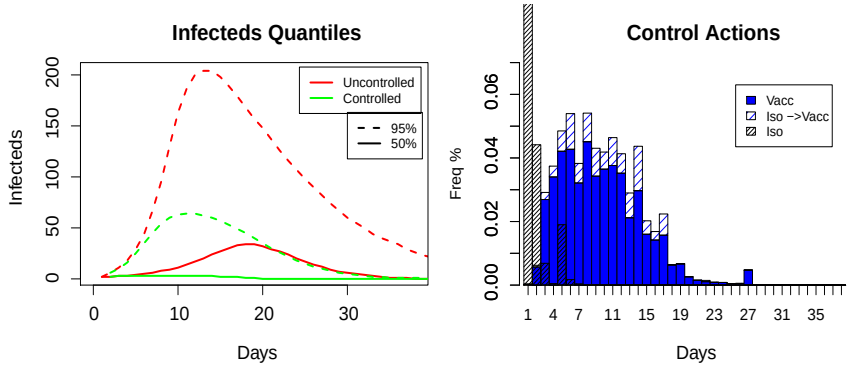

Figure 4: Effect of dynamic control policy versus no interventions. Left plot shows median (solid) and 95%-quantile (dashed) for dynamic policy (green) and no intervention policy (red). Right plot shows the frequency of new control actions as a function of time over 25,000 optimally controlled scenarios. We present two histograms, one for starting Isolation (black) and another for starting Vaccination (blue). The second histogram is further subdivided into two cases depending on whether the previous policy was Wait-and-see or Isolation.

The right plot of Figure 4 provides the timing of policy decisions under the computed control strategy. We plot the expected distribution of the interventions over time, given the priors at $t=0$. As expected, isolation is usually undertaken early on as a preventive step. In contrast, vaccination normally occurs during the second week when its cost declines sufficiently to be economical. While an uncontrolled outbreak often lasts well into the second month, with this dynamic strategy over 90% of outbreaks are completely extinguished by $t=30$. Overall, in these simulations, no intervention is taken 35% of the time, vaccination alone is applied in 39% of the scenarios, isolation alone in 16% of scenarios, and 10% of the time isolation is combined either sequentially or simultaneously with vaccination.

To provide yet another perspective on optimal policy, Figure 5 shows the distribution of dynamic controls as a function of true outbreak parameters. Namely, for each combination of $(\mathcal{R}_0, 1/\gamma)$ we run our adaptive algorithm 1000 times and tabulate the frequency of various policy choices encountered. Due to the stochastic fluctuations, the scenarios generated even with fixed (β, γ) can be quite different (see Figure 3); this is even more true in the adaptive case where early decisions will influence future outbreak evolution. As expected, for small $\mathcal{R}_0$, the wait-and-see and isolation strategies (chosen based on typical

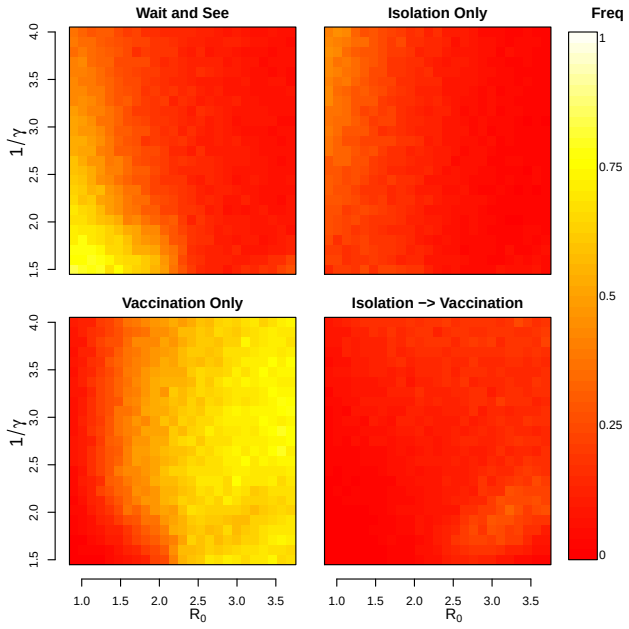

Figure 5: Policy implementation frequency as a function of the basic reproductive ratio $\mathcal{R}_0$ and average length of morbidity $1/\gamma$ in days for an infected individual.

duration of morbidity γ^{-1}) are sufficient to prevent a major outbreak and minimize costs. For moderate and large $\mathcal{R}_0$ vaccination is undertaken; for truly severe outbreaks that move quickly (lower right corner in Figure 5), the joint policy of isolation and vaccination (usually applied sequentially) is necessary.

The current example assumes morbidity costs are linear in the number of infecteds. Under such costs, the burden of severe outbreaks is small. In order to put more weight on preventing major outbreaks, we consider nonlinear morbidity costs. Namely, we take $c_1(I_t) = I_t + 0.1[\max(I_t - \bar{I}, 0)]^2$ which imposes a quadratic penalty whenever $I_t > \bar{I}$. As $\bar{I}$ is lowered, a more preventive policy should be undertaken to keep $c_1(I_t)$ small. Table 1 provides detail about expected cost and policy mix of our dynamic flu management for this nonlinear morbidity cost. For a linear cost structure ($\bar{I} = \infty$), vaccination is preferred over isolation. However, if the outbreak must be kept small, isolation becomes useful to forestall growth in I_t until vaccination cost drops enough. At the extreme, for $\bar{I} = 25$, it is optimal to preventively isolate all scenarios at $t = 0$.

Table 1: Dynamic policy mixes under different cost structures where $c_1(i) = i + 0.1[\max(i - \bar{I}, 0)]^2$ and $c_7(t_V) = c_7 e^{-(t_V - 1)/7}$. The table presents the frequency of each policy response given the cost structure (if both isolation and vaccination is applied for the same scenario, it is reported in both columns). The last two rows show the effect of vaccination costs on policy mix. We take $\bar{I} = \infty$ and vary the fixed cost $c_7(t)$ charged if vaccination is started at date t.

$\bar{I}$	c_7	Exp. Cost	Isolation	Vaccination
∞	2500	876	25%	49%
75	2500	901	8%	55%
50	2500	946	11%	51%
25	2500	1002	100%	44%
∞	2000	834	22%	55%
∞	3000	903	100%	28%

4.2 Comparison of boarding school policy strategies

To determine whether our sophisticated adaptive control strategy is worthwhile, we compare average costs under our strategy versus a number of other strategies under several different cost structures. In particular, we compare our dynamic policy to the policies of

- No control: wait-and-see always,
- Isolation: isolation always applied immediately at $t = 1$ and maintained throughout,
- Threshold: start isolation as soon as $I_t > 2$; start vaccination as soon as $I_t > 12$, and
- Fixed-date: an optimal one-time policy choice at fixed time $t = 7$ (similar to Merl et al. (2009)).

Table 2 presents a comparison of average cost under the quadratic cost structures. In all cases, the dynamic policy presented here lowers expected costs by at least 8%. As expected, with more nonlinear cost structures, the benefit of a dynamic policy is even more pronounced since the algorithm automatically fine-tunes to minimize the modified expected costs. A useful benchmark is the fixed-date policy, studied in Merl et al. (2009), which is computed using our Algorithm 1 assuming that control can only be applied at a single time point

but otherwise taking into account the updated posteriors. Thus, the difference between the fixed-date and dynamic policy showcases the importance of decision timing for flu outbreaks.

Table 2: Average cost comparison of the dynamic control strategy described here versus no control, isolation always, threshold, and fixed-date for various cost structures where $c_1(i) = i + 0.1[\max(i - \bar{I}, 0)]^2$. MC error of each average cost is about 0.5%.

$\bar{I}$	Wait-and-see	Isolation	Threshold	Fixed-date	Dynamic
∞	1030	1400	1091	942	876
75	3350	1440	1091	1183	886
50	4950	1515	1091	1378	902
25	7250	1666	1097	1667	931

The impressive savings shown in Table 2 are overestimated since we did not optimize the parameters of the alternative control strategies. Thus, the gain of our dynamic policy would be lessened if we were to find the optimal isolation start time, threshold values, or fixed date. However, in essence, this is exactly what our global algorithm does as it searches through all possible control policies, including those reference ones and automatically chooses the optimal strategy.

5 Discussion

We have presented a methodology to find optimal policy maps for dynamic risk management of flu outbreaks in a stochastic framework with parameter uncertainty. The methodology accounts both for stochastic interactions between individuals during outbreaks, as well as uncertainty about the outbreak parameters that are important for policy makers. The simulation-based control algorithm computes an approximately optimal adaptive and dynamic management strategy which creates a full policy map across all possible outbreak scenarios. This improves on existing literature that searches only through a small set of pre-determined vaccination strategies or uses point estimates of parameters rather than the full posteriors. Simultaneously, our use of regression Monte Carlo eschews the curse of dimensionality associated with classical dynamic programming in high dimensions (recall that our full state vector x_t is six-dimensional).

Parameter uncertainty is central to our approach and its effect on expected outbreak costs is two-fold. First, the expected number of infecteds is nonlinear in the outbreak parameters. Coupled with the typical convex cost structure, this implies that expected costs increase as a function of variance in parameter priors. Second, parameter uncertainty implies that suboptimal decisions will be undertaken due to incomplete information. For instance, an outbreak with moderate β and small γ (hence large $\mathcal{R}_0$) might initially "masquerade" as a mild epidemic since, initially, the number of infecteds will grow slowly. Thus, it might appear that isolation is a sufficient control strategy. As the outbreak begins in earnest, the intervention will be switched to vaccination. If the outbreak parameters were known from the outset, one could simply wait till the outbreak begins and immediately apply vaccination, saving isolation costs. Such unnecessary interventions will increase in frequency as uncertainty rises since uninformative parameter estimates will make it harder to distinguish outbreak properties.

Our algorithm can be easily extended to incorporate more sophisticated epidemic models beyond the general stochastic SIR presented above. We may allow time-inhomogeneous reaction rates to account for seasonality of flu. We can also straightforwardly add further individual classifications, including Exposed, Diseased, and other groups. Finally, we can also consider more complicated models, e.g. household-level data. This would replace the two-dimensional state vector (S_t, I_t) with a more complete description $(S_{t,h}, I_{t,h})$ where $(S_{t,h}, I_{t,h})$ denotes the number of susceptibles and infecteds at time t in household h. Such structure would clearly increase the dimensionality of the state vector, however one expects that a few summary statistics, such as percentage of households with no infecteds, can be identified that are key for policy making.

We can also include more complicated interventions, such as treatments that speed up recovery, on-the-spot vaccination that instantaneously cuts down the number of susceptibles, or conversely delayed vaccination effects. A particularly interesting possibility is to allow for information-gathering through e.g. genetic testing of the pathogen, which would reduce the uncertainty about outbreak parameters and mitigate the associated costs discussed above. Such a control would act directly on the posteriors Z_t, rather than on the $S \to I \to R$ transition rates. All such modifications would simply increase/modify the dimension and effect of the control vector ϕ. In the same spirit, we can allow for variability in control effects, so that policies have a random or unknown impact on outbreak progression which itself requires estimation. Finally, after adjusting the set of possible controls and state transitions other diseases beyond human influenza may be considered.

In the model presented, we had complete observations of the state vector which is realistic for outbreaks in small, closed communities where medical officers can monitor each individual. In larger groups, numbers of susceptibles/infecteds will be unknown and will need to be inferred dynamically as the outbreak progresses. Such incomplete observation models require more sophisticated statistical estimation techniques and will be dealt with in a separate forthcoming work.

A Simulating Epidemic Spread

Our control algorithm is based on optimization over a large set of outbreak scenarios that capture potential evolution of the outbreak state in the future while subject to dynamic management. In this section we briefly describe how such outbreak scenarios can be efficiently simulated on a computer.

The state vector (S_t, I_t) forms a continuous time Markov chain according to the transition probabilities of equation (1). Appealing to the transition formulation, simulation of outbreak spread can be achieved through any of the variants of the Gillespie stochastic simulation algorithm (SSA). We describe the simplest SSA approach here (Gillespie, 1977) and refer readers to Gillespie (2007) for possible speed-ups and approximations.

Let us denote by t_k, $k = 0, 1, 2, \ldots$ the transition times of the (S_t, I_t) chain. The SSA is initialized with state (S_0, I_0) and time $t = t_0 = 0$. If the system is in state (S_{t_k}, I_{t_k}) at time t_k, SSA simulates a transition by first sampling a time to the next transition $t_{k+1} - t_k$ which has distribution $Exp\big(k_I^{\phi^I_{t_k}} \beta I_{t_k} S_{t_k}/N + \gamma I_{t_k} + \phi^V_{t_k} k_V S_{t_k}\big)$ where $Exp(\lambda)$ indicates an exponential random variable with mean $1/\lambda$. Then a transition indicator $j \in \{1, 2, 3\}$ is sampled with probability proportional to $k_I^{\phi^I_{t_k}} \beta I_{t_k} S_{t_k}/N$ for $j = 1$, γI_{t_k} for $j = 2$, and $\phi^V_{t_k} k_V S_{t_k}$ for $j = 3$. Time is incremented to $t = t_{k+1}$ and the state is updated depending on the transition sampled, i.e. $S_{t_{k+1}} = S_{t_k} - 1$ if $j \in \{1, 3\}$ while $S_{t_{k+1}} = S_{t_k}$ if $j = 2$, and $I_{t_{k+1}} = I_{t_k} + 1$ if $j = 1$, $I_{t_{k+1}} = I_{t_k} - 1$ if $j = 2$, and $I_{t_{k+1}} = I_{t_k}$ if $j = 3$.

Due to the memory-less property of exponential distributions, the mean of the exponential can be recalculated at time t_{k+1} and the time to the next transition sampled along with a transition indicator from the updated probabilities. Continuing this procedure forward, a simulation of the system from time $t = 0$ to any time T can be obtained. The outbreak ends when $I_t = 0$ at which time the rate of the exponential is zero indicating that no further transitions need to be simulated.

B Algorithms

B.1 Sequential flu management algorithm

Algorithm 1 creates the (approximately) optimal policy map $\Phi(t, x_t, \phi_{t-1})$ for all time points $t \in \{0, 1, \ldots, T-1\}$. As input, the algorithm requires 1) the number of simulations K to perform at all time points, 2) a set of outbreak statistics $B(x_t)$ of dimension r to use in the regression step, and 3) a default policy map Φ^0 for initial simulation. The algorithm returns the regression coefficients α_t^ϕ. The total computational cost of this algorithm is $T^2 \cdot K \cdot \max_\phi |F(\phi)|^2$. This is also how many times Algorithm 2 is run and how many times we need to perform the stochastic simulation of $(S_{0:T}, I_{0:T})$.

Policy maps are created in real-time using the regression coefficients α^ϕ. At time t with state X_t and previously implemented policy ϕ_{t-1}, the policy map is

$$\Phi(t, X_t, \phi_{t-1}) = \underset{\phi \in F(\phi_{t-1})}{\operatorname{argmin}} \; (\alpha_t^\phi)' B(X_t)$$

and determines the optimal policy ϕ_t to implement.

Algorithm 1 Flu management policy map creation

for $k \in \{1, 2, \ldots, K\}$ **do**
  Simulate $x_{1:T}^{(k)}$ using Algorithm 2 and default policy map Φ^0
end for
for $t = (T-1), \ldots, 1, 0$ **do**
  Evaluate the outbreak statistics $B(x_t^{(k)})$.
  for each policy ϕ **do**
    for $k \in \{1, 2, \ldots, K\}$ **do**
      Obtain a realized cost $v_t^{k,\phi}$ by simulating $\tilde{x}_{t:T}^{(k,\phi)}$ starting from $\tilde{x}_t^{(k)} = x_t^{(k)}$ using Algorithm 2 and implementing ϕ at decision time t.
    end for
    Compute the regression coefficients

$$\alpha_t^\phi = \underset{\alpha \in \mathbb{R}^r}{\operatorname{argmin}} \sum_{k=1}^{K} \left[v_t^{k,\phi} - \alpha' B\left(x_t^{(k)}\right) \right]^2.$$

  end for
end for
return Return regression coefficients $\alpha_{0:T}^\phi$ for all ϕ.

B.2 Forward simulation and associated costs

Algorithm 2 is the workhorse of our approach and consists of simulating a future controlled outbreak scenario from time t until time T and calculating the associated cost. The required input is 1) a policy map $\Phi(s, x_s, \phi_{s-1})$ for $s \in \{t, t+1, \ldots, T-1\}$ that determines the policy choice for any given outbreak realization and current control and 2) initial conditions ϕ_t and $x_t = (s_t, i_t, z_t)$ where z_t refers to the current shape and rate parameters determining the time-t posteriors for β, γ. The algorithm returns realized outbreak scenarios $X_{t:T}$ and the realized costs v_t.

Algorithm 2 Simulate one outbreak and calculate costs v_t based on policy map.

Initialize $v_t \leftarrow 0$ // Realized costs from t to T
for $s = t, \ldots, T-1$ **do**
  Set $\phi_s \leftarrow \Phi(s, x_s, \phi_{s-1})$.
  Sample β, γ from $p(\beta, \gamma | z_s)$.
  Simulate outbreak trajectory from $p(S_{(s,s+1]}, I_{(s,s+1]} | s_s, i_s, \beta, \gamma, \phi_s)$ using SSA of Appendix A.
  Update hyperparameters z_{s+1} based on $s_{(s,s+1]}, i_{(s,s+1]}$ and z_s according to equation (3).
  Update simulated costs $v_t \leftarrow v_t + C_s^{s+1}(\phi_s)$ according to equation (4).
end for
return Simulated outbreak scenario $X_{t:T}$ and cumulative cost v_t.

References

Andersson, H. and T. Britton (2000): *Stochastic epidemic models and their statistical analysis*, *Lecture Notes in Statistics*, volume 151, New York: Springer-Verlag.

Bailey, N. (1975): *The Mathematical Theory of Infectious Diseases.*, Griffin, London.

Ball, F. and O. Lyne (2002): "Optimal vaccination policies for stochastic epidemics among a population of households," *Mathematical Biosciences*, 177-178, 333–354.

Ball, F. and P. Neal (2002): "A general model for stochastic SIR epidemics with two levels of mixing," *Mathematical Biosciences*, 180, 73–102.

Bertsekas, D. P. (2005): *Dynamic programming and optimal control. Vol. I*, Athena Scientific, Belmont, MA, third edition.

Cai, H. and X. Luo (1994): "Stochastic control of an epidemic process," *International Journal of Systems Science*, 25, 821–828.

Cintron-Arias, A., C. Castillo-Chavez, L. Bettencourt, A. Lloyd, and H. Banks (2009): "The estimation of the effective reproductive number from disease outbreak data," *Mathematical Biosciences and Engineering*, 6 (2), 261–282.

Clancy, D. and N. Green (2007): "Optimal intervention for an epidemic model under parameter uncertainty," *Mathematical Biosciences*, 205, 297–314.

Dukic, V., H. Lopes, and N. Polson (2010): "Tracking flu epidemic using Google Flu Trends and particle learning," Working paper.

Egloff, D. (2005): "Monte Carlo algorithms for optimal stopping and statistical learning," *Annals of Applied Probability*, 15, 1396–1432.

Elderd, B., V. Dukic, and G. Dwyer (2006): "Uncertainty in prediction of disease spread and public health responses to bioterrorism and emerging diseases," *PNAS*, 103, 15693–15697.

Fleming, W. H. and H. M. Soner (1993): *Controlled Markov processes and viscosity solutions*, *Applications of Mathematics (New York)*, volume 25, New York: Springer-Verlag.

Germann, T. C., K. Kadau, I. M. Longini, and C. A. Macken (2006): "Mitigation strategies for pandemic influenza in the United States," *PNAS*, 103 (15), 5935–5940.

Gillespie, D. T. (1977): "Exact stochastic simulation of coupled chemical reactions," *Journal of Physical Chemistry*, 81, 2340–2361.

Gillespie, D. T. (2007): "Stochastic simulation of chemical kinetics," *Annual Reviews in Physical Chemistry*, 58, 35–55.

Halloran, M. E., N. M. Ferguson, S. Eubank, I. M. Longini, D. A. T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T. C. Germann, D. Wagener, R. Beckman, K. Kadau, C. Barrett, C. A. Macken, D. S. Burke, and P. Cooley (2008): "Modeling targeted layered containment of an influenza pandemic in the United States," *PNAS*, 105 (12), 4639–4644.

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

Longstaff, F. and E. Schwartz (2001): "Valuing American options by simulations: a simple least squares approach," *Rev. Finan. Studies*, 14, 113–148.

Ludkovski, M. (2005): *Optimal Switching with Application to Energy Tolling Agreements*, Ph.D. thesis, Princeton University.

Ludkovski, M. (2009): "A simulation approach to optimal stopping under partial observations," *Stochastic Processes and Applications*, 119(12), 4061–4087.

Martínez-Beneito, D. Conesa, A. López-Quílez, and A. López-Maside (2008): "Bayesian Markov switching models for the early detection of influenza epidemics," *Statistics in Medicine*, 27, 4455–4468.

Meltzer, M. I., N. J. Cox, and K. Fukuda (1999): "The economic impact of pandemic influenza in the United States: Priorities for intervention," *Emerging Infectious Diseases*, 5 (5), 659–671.

Merl, D., L. Johnson, R. Gramacy, and M. Mangel (2009): "A statistical framework for the adaptive management of epidemiological interventions," *PLoS ONE*, 4(6), e5087.

Murray, J. (2002): *Mathematical biology. I. An Introduction*, New York: Springer-Verlag, 3rd edition.

Niemi, J. (2009): *Bayesian analysis and computational methods for dynamic modeling*, Ph.D. thesis, Duke University.

O'Neill, P. D. (2002): "A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods," *Mathematical Biosciences*, 180, 103–114.

Patel, R., I. Longini, and M. Halloran (2005): "Finding optimal vaccination strategies for pandemic inuenza using genetic algorithms," *Journal of Theoretical Biology*, 234, 201212.

R Development Core Team (2009): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`, ISBN 3-900051-07-0.

Tanner, M. W., L. Sattenspiel, and L. Ntaimo (2008): "Finding optimal vaccination strategies under parameter uncertainty using stochastic programming," *Mathematical Biosciences*, 215, 144–151.

Tennenbaum, S. (2008): "Simple criteria for finding (nearly) optimal vaccination strategies," *Journal of Theoretical Biology*, 250, 673–683.

Wilkinson, D. J. (2006): *Stochastic Modelling for Systems Biology*, London: Chapman & Hall/CRC.

Yang, Y., J. D. Sugimoto, M. E. Halloran, N. E. Basta, D. L. Chao, L. Matrajt, G. Potter, E. Kenah, and I. M. Longini (2009): "The transmissibility and control of pandemic influenza A (H1N1) virus," *Science*, 326, 729.