amei: an **R** package for the Adaptive Management of Epidemiological Interventions

Daniel Merl
University of California, Santa Cruz

Leah R. Johnson
University of California, Santa Barbara

Robert B. Gramacy
University of Cambridge

Marc Mangel
University of California, Santa Cruz

Abstract

The **amei** package for R is a tool that provides a flexible statistical framework for generating optimal epidemiological interventions that are designed to minimize the total expected cost of an emerging epidemic. Uncertainty regarding the underlying disease parameters is propagated through to the decision process via Bayesian posterior inference. The strategies produced through this framework are adaptive: vaccination schedules are iteratively adjusted to reflect the anticipated trajectory of the epidemic given the current population state and updated parameter estimates. This document, which is (slightly) modified version of Merl *et al.* (2010), published in the *Journal of Statistical Software*, briefly covers the background and methodology underpinning the implementation provided by the package and contains extensive examples showing the functions and methods in action.

Keywords: SIR model, Bayesian inference, optimal decisions, Markov chain Monte Carlo.

1. Introduction

The goal of the study of infectious diseases is to better understand how infections are spread and maintained, and ultimately to find ways to control the spread of a disease. The most common methods for intervening in the spread of an infectious disease are to either remove susceptible individuals or apply treatment to infected individuals. For instance, the susceptible population may be culled, as in the case of foot-and-mouth disease (Tildesley *et al.* 2006; Enserink 2001), or the infected population may be quarantined, as in the case of SARS (Lloyd-Smith *et al.* 2003). Most commonly, susceptibles may be vaccinated, as in the case with influenza or smallpox (Ferguson *et al.* 2003; Halloran *et al.* 2002).

Each of these actions incurs a quantifiable epidemiological cost. For culling, the cost is an additional number of deaths; for quarantine, the cost is likely to be measured in monetary units rather than lost lives; for vaccination, the cost may be measured in both monetary units as well the number of additional vaccine–induced infections; for medical treatment, the cost is again monetary. Furthermore, the costs associated with each action can depend upon the state of the disease within the population of interest. This raises the question of how to find optimal epidemiological interventions in a manner that adaptively depends on the state of

the epidemic.

Most existing methods for finding optimal intervention strategies are concerned with the situation of pre-emptive intervention which is assumed to be completed before the onset of the epidemic (for instance see Ball and Lyne 2002; Patel *et al.* 2005; Tildesley *et al.* 2006). In this kind of situation, there is no reason to consider sequentially updated (i.e. adaptive) interventions: as soon as the intervention policy is triggered, the epidemic threat will be eradicated. However, in most scenarios, total and instantaneous intervention will not be an implementable strategy. Moreover, these methods usually evaluate the cost of the epidemic in a manner that assumes no uncertainty in key model parameters, including transmission rate, recovery rate, and others (however, see Elderd *et al.* 2006 and our brief discussion in Section 5).

Here we introduce **amei**, an R package (R Development Core Team 2009) that implements the statistical framework introduced by Merl *et al.* (2009). This framework allows one to respond to an emerging epidemic while simultaneously learning about its evolutionary dynamics. We consider vaccination strategies defined by a fraction of the current susceptible population to be targeted for vaccination, and a threshold number of susceptibles below which the vaccination campaign is called off. We couple the evaluation of optimal, adaptive, intervention strategies with Bayesian procedures for performing on-line estimation of the parameters of the underlying epidemic model, thereby propagating parameter uncertainty through to policy decisions. We demonstrate the advantages of adaptive intervention via the functions provided by the package using simulations modeled after an influenza outbreak at a British boarding school described by Murray (2002). We compare the distribution of costs arising from epidemiological intervention under the adaptive policies to those arising from non-adaptive policies (i.e. policies not dependent on the state of the epidemic and/or not reflecting parameter uncertainty), and find that the adaptive policies result in low total costs, efficient use of available resources, and are robust to model misspecification.

This tutorial is comprised of three main sections. The mathematical specification of the Bayesian models used for inference and the Monte Carlo methods for constructing optimal vaccination strategies (both static or adaptive and on–line) are contained in Section 2. In Section 3, the functions and methods implemented in the package are illustrated by following a single, detailed, example, whose results were first reported by Merl *et al.* (2009). Section 4 contains a second example with an alternative transmission model, thereby illustrating the additional benefits of the approach and features of the package. The paper concludes in Section 5 with a discussion of the methodology and related work highlighting other freely available software with comparisons and contrasts. Miscellaneous details on implementation, etc., are provided in an appendix.

2. Methods

2.1. SIR Model

In **amei**, we consider a stochastic version of a standard Susceptible–Infected–Removed (SIR) model (Anderson and May 1991; Hetchote 2000) with permanent immunity and with mortality. In this model, the dynamic variables at time t are the number of susceptible individuals, $S(t)$; the number of infected individuals, $I(t)$; the number of recovered individuals, $R(t)$; and

the number of removed/dead individuals, $D(t)$. In a standard SIR model we would typically lump all of the Recovered and Dead individuals into a single removed class. However, in controlling an epidemic we expect there to be differential costs associated with deaths versus other types of removal (see below), so we count the numbers of deaths separately. We assume the population is closed to immigration or emigration so that $S(t) + I(t) + R(t) + D(t) = N$, where N is constant.

We adopt a flexible negative binomial form for the transmission function (Godfray and Hassell 1989; Mangel 2006; McCallum *et al.* 2001). Under this assumption, the SIR model is described by the following system of differential equations (Hetchote 2000; McCallum *et al.* 2001):

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -kS\ln\left(1 + \frac{bI}{k}\right) \tag{1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = kS\ln\left(1 + \frac{bI}{k}\right) - (\nu + \mu)I$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \nu I \tag{2}$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \mu I \tag{3}$$

The model parameters are: the transmission rate b; the overdispersion (or "clumpiness") parameter k; the death rate μ; and the rate of recovery to the immune class ν. The negative binomial distribution can be interpreted as a compound stochastic process in which encounters between infected and susceptible individuals occur randomly (i.e. according to a Poisson process) such that the encounter rate varies according to a gamma distribution with coefficient of variation $k^{-1/2}$. Thus, via k, the negative binomial transmission can account for social interactions and/or network factors in disease transmission, without requiring explicit characterization of the population structure.

This SIR formulation leads to a natural discrete time approximation for the numbers of infections ($\tilde{I}$), recoveries ($\tilde{R}$), and deaths ($\tilde{D}$) arising in the unit time interval from t to $t + 1$. Assuming the total number of infected individuals, I, is approximately constant and integrating Eq. (1) over a unit time interval gives

$$S(t + 1) = S(t)\left[\frac{k}{k + bI(t)}\right]^k, \tag{4}$$

so that the fraction of susceptible individuals surviving throughout a unit time interval is $\left[\frac{k}{k+bI(t)}\right]^k$. When viewed as a discrete time stochastic process, where the mean number of remaining susceptible individuals is given by Eq. (4), the mean number of new infections occurring between time t and $t + 1$ is

$$S(t)\left(1 - \left[\frac{k}{k + bI(t)}\right]^k\right).$$

Therefore, if $S(t) = s$ and $I(t) = i$, we may sensibly take the new infections $\tilde{I}$ at time $t + 1$ to follow

$$\tilde{I}|s,i \sim \mathrm{Bin}(s, p_i(i, b, k)), \qquad \text{where} \qquad p_i(i, b, k) = 1 - \left(\frac{k}{k + bi}\right)^k, \tag{5}$$

and $\mathrm{Bin}(n, \pi)$ is the binomial distribution with size n and success probability π. Similarly, by integrating Eqs. (2–3), the numbers of recoveries and deaths occurring between time t and $t + 1$ can be described by

$$
\begin{align}
\tilde{R}|i &\sim \mathrm{Bin}(i, p_r) \tag{6} \\
\tilde{D}|i, \tilde{r} &\sim \mathrm{Bin}(i - \tilde{r}, p_d) \tag{7}
\end{align}
$$

where $p_r = 1 - e^{-\nu}$ and $p_d = 1 - e^{-\mu}$. The forward dynamics for the total numbers of susceptible and infected individuals are therefore

$$
\begin{align}
S(t + 1) =& S(t) - \tilde{I}|s, i \\
I(t + 1) =& I(t) + (\tilde{I}|s, i) - (\tilde{R}|i) - (\tilde{D}|i, \tilde{r} + \tilde{i}).
\end{align}
$$

Here the lower case symbols $\{s, i, r\}$ denote the realized value of the associated capital letter random variable. In this discrete time approximation we have assumed a particular ordering of events, namely that recoveries occur first, followed by deaths from among those infected individuals who did not recover, followed by new infections. Simulation studies indicated that these assumptions, as well as other possible orderings, resulted in system dynamics that were equal in expectation to the deterministic solutions to the continuous time SIR model.

2.2. Online Parameter Estimation

A primary task of **amei** is to perform inference on the SIR model parameters, conditional on sequential observations of the numbers of susceptible, infected, and recovered individuals in the population. Given the discrete time approximation in the previous section, it is possible to do this online (i.e. as the epidemic progresses) via straightforward parametric Bayesian methods. In particular, we use Markov Chain Monte Carlo (MCMC) (e.g. Gamerman and Lopes 2006) to learn about the posterior distributions of b, k, ν, and μ conditioned on the evolution of the epidemic observed so far. The likelihood is given recursively in Eq. (5–7). Let $\tilde{i}_t = S(t - 1) - S(t)$ be the number of new infecteds at time T, and similarly for the newly recovered and dead individuals $\tilde{r}_t$ and $\tilde{d}_t$ so that $\tilde{r}_t + \tilde{d}_t \leq I(t - 1)$. Then, the likelihood of the data observed up to time T is given by

$$
\begin{align}
&\prod_{t=1}^{T} \mathrm{Bin}(\tilde{i}_t \,|\, S(t - 1), p_i(I(t - 1), b, k)) \\
&\times \prod_{t=1}^{T} \mathrm{Bin}(\tilde{r}_t | p_r) \times \prod_{t=1}^{T} \mathrm{Bin}(\tilde{d}_t | I(t) - \tilde{r}_t, p_d).
\end{align}
$$

Observe that it consists of three mutually independent components when conditioning on the course of the epidemic.

Conditional conjugacy can be exploited for ν and μ via beta priors for p_r and p_d. A $\mathrm{Beta}(\alpha_r, \beta_r)$ prior for p_r implies that

$$
p(\nu) = (1 - e^{-v})^{\alpha_r - 1} e^{-v\beta_r}. \tag{8}
$$

Conjugate updating leads to the posterior conditional

$$
p_r | \cdots \sim \mathrm{Beta}\left(\alpha_r + \textstyle\sum_{t=1}^{T} \tilde{r}_t, \beta_r + \sum_{t=1}^{T} I(t) - \tilde{r}_t\right). \tag{9}
$$

The form of the conditional posterior for ν is similar to Eq. (8) and can be simulated by first drawing p_r via Eq. (9) and then applying the inverse transformation $\nu = -\log(1 - p_d)$. Sampling for μ proceeds similarly with

$$p_d| \cdots \sim \text{Beta}\left(\alpha_d + \sum_{t=1}^{T} \tilde{d}_t, \beta_d + \sum_{t=1}^{T} I(t) - \tilde{r}_t - \tilde{d}_t\right). \tag{10}$$

Thus it is possible to take Gibbs samples for ν and μ so long as appropriate hyperparameters $\alpha_r, \beta_r, \alpha_d, \beta_d$ can be found to represent our prior beliefs. In ignorance we simply set these to unity by default, leading to a uniform prior on p_r and p_d. The user is free to specify his/her own prior parameterization in the package.

Obtaining samples for b and k requires the Metropolis–Hastings algorithm. Our prior beliefs can be encoded with gamma distributions, and conditional on a previous sample (b, k) the next sample (b', k') can be obtained by Metropolis–within–Gibbs steps using:

$$p(b'|k, \dots) \tag{11}$$

$$\propto \Gamma[b'|\alpha_b, \beta_b] \prod_{t=1}^{T} \text{Bin}(\tilde{i}_t \,|\, S(t-1), p_i(I(t-1), b', k)), \qquad \text{and}$$

$$p(k'|b', \dots) \tag{12}$$

$$\propto \Gamma[k'|\alpha_k, \beta_k] \prod_{t=1}^{T} \text{Bin}(\tilde{i}_t \,|\, S(t-1), p_i(I(t-1), b', k')).$$

The default values for the prior parameterization set in the package are $(\alpha_b, \beta_b) = (\alpha_k, \beta_k) = (1, 3)$ which (though seemingly informative at first glance) turns out to be uninformative on the scale of the support of the posterior. As before, these can easily be changed by the user. We use random walk uniform proposals on the positive real line, i.e. $b' \sim U[3b/4, 4b/3]$, which gives reasonably good mixing in the Markov chain.

The presence of a vaccination strategy (described in the next section) necessitates a simple change to the above equations. If $0 \leq v_t \leq S(t-1)$ is the number of susceptibles which have been vaccinated, then we simply replace $S(t-1)$ with $S(t-1) - v_t$ so that $\tilde{i}_t = S(t-1) - v_t - S(t)$.

2.3. Optimal Vaccination Strategies: Fixed and Adaptive

Having obtained estimates of the SIR model parameters at some time during the epidemic, we next want find the best way to intervene in the spread of the epidemic. The first step is to formalize the notion of the "best" intervention strategy. This requires a specification of the costs of various actions, such as vaccination, versus the cost of allowing the epidemic to spread in an uncontrolled fashion. In **amei**, we formulate the total expected cost of the epidemic in terms of the underlying costs associated with maintaining infected individuals until recovery, suffering death, and administering vaccinations. We assume costs associated with each action can be specified in some common currency, such as monetary costs, or probabilities of mortality (such that the cost to be minimized is the expected numbers of deaths).

We formulate our vaccination strategies in terms of a policy where a fraction, α, of susceptibles are prevented from risk of infection by moving them directly into an immune/recovered class, such as by perfect vaccination, until the number of individuals that are still susceptible drop

below a threshold, γ, and vaccination is discontinued. Let $c_1(\alpha, \gamma, s)$ denote the cost associated with the vaccination strategy (α, γ) when $S(t) = s$. If c_v denotes the cost per vaccine unit, then

$$c_1(\alpha, \gamma, s) = \begin{cases} c_v \alpha s & \text{if } s > \gamma \\ 0 & \text{if } s \leq \gamma. \end{cases}$$

Thus the number if individuals vaccinated at time t is $v_t = \alpha s$. Let $c_2(i)$ denote the cost component that depends on the number of infections in the population, $I(t) = i$. This component includes the costs associated with maintaining the non-recovered infected individuals and costs associated with deaths, as in

$$c_2(i) = c_t i + c_d \tilde{d},$$

where c_t is the cost per treatment/maintenance of a non-removed infected individual, and c_d is the cost per death.

Assuming the initial epidemiological state is $S(0) = s_0$, $I(0) = i_0$, the expected total cost of the epidemic under intervention strategy (α, γ) can be expressed recursively as

$$\mathsf{E}\{C_0\} = c_1(\alpha, \gamma, s_0) + c_2(i_0) + \mathsf{E}\{C_1\}, \tag{13}$$

where $\mathsf{E}\{C_t\}$ denotes the expected cost accumulated from time t onwards, terminating at some arbitrary time horizon. The optimal intervention strategy (α, γ) is the one that minimizes the total accumulated cost over the course of the epidemic. Two methods for calculating such strategies are as follows.

The first case we are interested in is when the parameters of the SIR model are known, and we wish to calculate the single best intervention strategy (α, γ). The total expected cost depends on the parameter values and the initial epidemiological state (s_0, i_0). Thus, conditional on the parameter values, Monte Carlo simulation of the disease dynamics specified in Section 2.1 can be used to search over values of α and γ in order to find the combination that minimizes $\mathsf{E}\{C_0\}$. For each combination of α and γ considered, we conduct n stochastic simulations of the outbreak in order to estimate the mean cost associated with the intervention (α, γ). The strategy producing the lowest mean cost is defined to be the optimal intervention. Typically we discretize and create a grid of admissible α and γ settings. In the examples in Section 3 we allow α to range from 0 to 1 in increments of 0.1, and γ to range from from 2 to $s_0 - s_0/10$ in increments of $s_0/10$, i.e. taking 10 steps.[1]

In the second case, we want to calculate an adaptive strategy that sequentially updates the best strategy (α, γ) as we learn more about the epidemic model parameters. As above, the expected cost surface associated with a given set of parameter values (as obtained by MCMC, described above), can be explored using standard Monte Carlo methods. At each time step, MCMC is used to produce samples from the current posterior distribution on model parameters conditional on all data observed up to that point. These samples are used to calculate the optimal vaccination strategy from that time point onward, as outlined as above. After calculating the optimal strategy associated with each set of posterior samples, the adaptive strategy to be implemented at that time step is defined to be the strategy that most frequently minimizes the expected cost over all samples.

---

[1] We generally do not include $\gamma = s_0$ in the grid since this policy (coupled with any α) is equivalent to $\alpha = 0$ for any γ.

3. An illustrative example

In this section we demonstrate the advantages of adaptive intervention using simulations modeled after an influenza outbreak at a British boarding school described by Murray (2002). We shall compare the distribution of costs arising from epidemiological intervention under the adaptive policies to those arising from non-adaptive policies. The epidemic conforms to many standard assumptions of SIR models: a population essentially closed to immigration and emigration; includes recovery and immunity; and has near homogeneous mixing of susceptibles and infecteds. The epidemic was traced back to a single infected student out of a population of 763 individuals.

For reproducibility of the results in this section we have set the random seed as follows.

```
R> set.seed(12345)
```

The following sections tacitly use similarly reset seeds.

We begin by exploring the behavior of the SIR model, described in Section 2.1, without any intervention. The first step is to set the relevant parameters that are necessary for simulating the epidemic. These consist of: the "true" underlying parameters for the SIR model upon which our costs will be evaluated; the initial condition of the population at the beginning of the epidemic; and the relative costs of infections, deaths, and vaccinations.

```
R> tp <- list(b = 0.00218, k = 10, nu = 0.4, mu = 0)
R> init <- list(S0 = 762, I0 = 1, R0 = 0, D0 = 0)
R> costs <- list(vac = 2, death = 4, infect = 1)
```

Murray (2002) provides estimates of the transmission rate (b) and recovery rate (ν), which we use in the `tp` "true parameterization" set above. We set the death rate (μ) to zero since there are no deaths in this epidemic. Finally, we set the negative binomial dispersion parameter (k) to be large to reflect the homogeneous mixing of the population. The `costs`, chosen purely for illustrative purposes, describe the unit cost (or loss) for a single vaccination or death, relative to the daily cost of maintaining an infected individual.

As a baseline, we are interested in the costs of a no–vaccination policy on epidemic trajectories with the above parameters. To explore this, we set the vaccination policy used in the simulation to zero.

```
R> vac <- list(frac = 0, stop = 0)
```
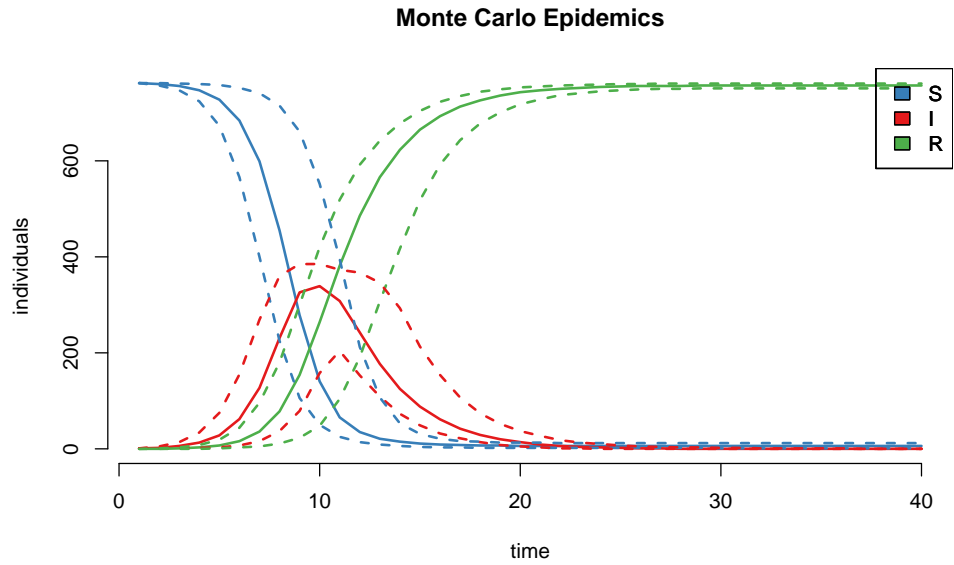
We are now ready to run the Monte Carlo experiment. The function `MCepi` can be used to simulate the population and cost trajectories for the experiment. This function simulates the stochastic SIR model under a given vaccination strategy. For each realization of the epidemic progression the function calculates the cost over time for the epidemic. This is repeated many times[2], and the mean trajectories for both the populations and the costs, as well as the 5th and 95th quantiles, are recorded.

```
R> init.MCepi <- MCepi(init, tp, vac, costs)
```

---

[2]The number of MC simulations can be varied with the `MCreps` argument to `MCepi`. The default value of `MCreps` is 1000.

Now we can simply plot the results to look at the distribution of the susceptible, infected, and recovered individuals in the population as the epidemic progresses (*top* panel of Figure 1), as well as the distribution of costs over time (*bottom* panel of Figure 1).

```
R> plot(init.MCepi)
```

Monte Carlo Epidemics


```
R> plot(init.MCepi, type = "costs")
```

Monte Carlo Costs

Figure 1: Monte Carlo simulated epidemic trajectories (*top*) for the numbers of susceptible, infected, and recovered individuals, and the associated cost(s) (*bottom*) with a *null* vaccination strategy. (2.5,50,97.5%) quantiles are shown.

Next we examine the dynamics of the system when we have perfect information and use a fixed vaccination policy. First construct a grid of admissible policies. In this case, we shall consider strategies that: vaccinate a fixed proportion of the susceptible population, from 0 to 100%, in steps of 10%; and that stop vaccinating susceptibles when the remaining susceptible population falls below some threshold between 2 and the initial susceptible population minus 75, in increments of 75, as explained in Section 2.3.

```
R> vacgrid <- list(fracs = seq(0, 1, 0.1), stops = seq(2,
+     init$S0 - 75, 75))
```

Once this grid has been initialized, we can run the Monte Carlo experiment, using the function `optvac`, which finds the vaccination policy that minimizes the total cost of the (stochastic) epidemic. This is done by simulating the epidemic forward under the known, `true`, parameterization and calculating the cost of each vaccination strategy.

```
R> out.optvac <- optvac(init, tp, vacgrid, costs)
```

This function outputs the costs for each of the possible vaccination strategies. The best and worst policies can be obtained as follows:

```
R> best <- getpolicy(out.optvac)
R> worst <- getpolicy(out.optvac, which = "worst")
R> rbind(best, worst)

      row col frac stop cost
best   10   3  0.9  152 1657
worst   1   1  0.0    2 2297
```

The same information can be obtained via the generic `print` and `summary` commands, which will be shown later. We can also plot the cost surface over the space of possible vaccination strategies. This takes the form of a heat plot, where lower cost policies appear in deep red, and high cost policies appear in white (Figure 2).

Given the calculated optimal policy, we can explore the effects of the vaccination strategy on the progression of the epidemic together with the trajectory of costs under this strategy. We do this by again simulating the epidemic dynamics using the function `MCepi`, however this time we include the best fixed vaccination policy.

```
R> vac.opt <- best[3:4]
R> opt.MCepi <- MCepi(init, tp, vac.opt, costs)
```

Figure 3 summarizes the trajectories (*top*) of the epidemic under the optimal vaccination strategy over time, and the corresponding costs (*bottom*). By default we assume a fixed lag of 7 time steps from when the first infection appears to when the first intervention can take place[3]. This is apparent in Figure 3, where there is a sharp transition from day 7 when the vaccinations begin, and the susceptible population drops dramatically. The number of vaccinated individuals can be added to the plot by specifying the argument `showv = TRUE`, however we omit this here to reduce clutter in the figure. Information on the distribution of the number of vaccine units dispensed can be extracted as follows.

---

[3]This can be varied with the `start` argument to `MCepi`.
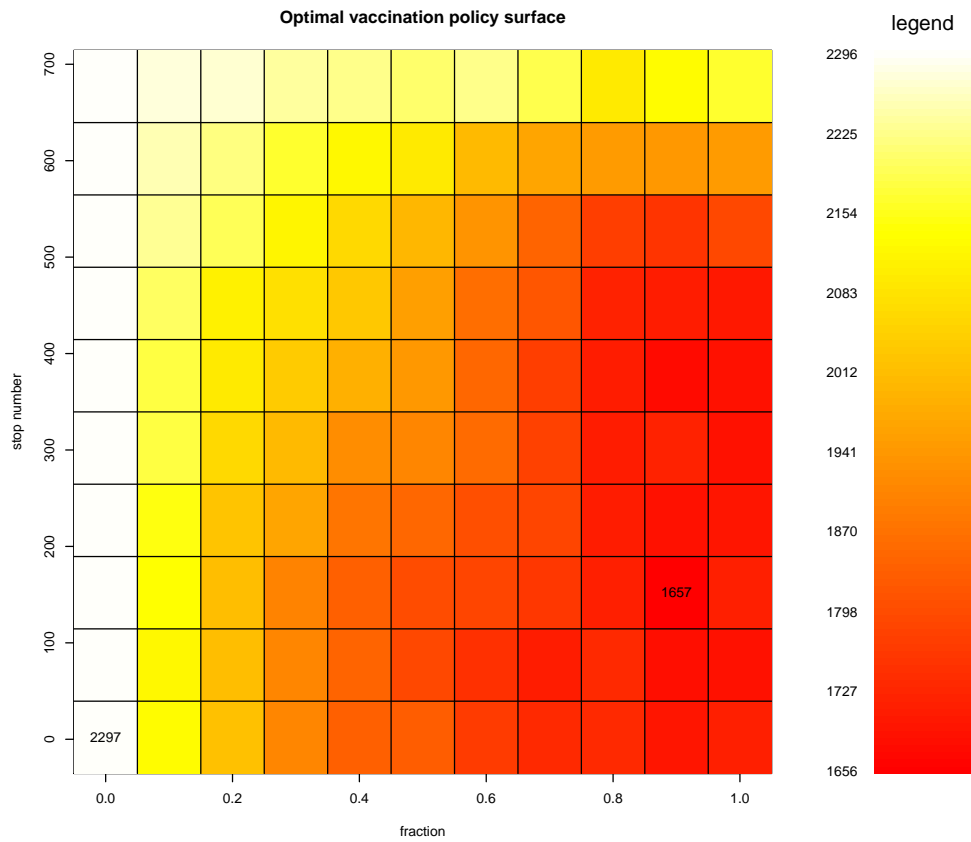
```
R> plot(out.optvac)
```


Figure 2: Heatmap depicting the expected cost surface associated with variable stop time vaccination strategies based on the true parameter values. The minimum expected cost (1657) is achieved by a strategy of vaccinating 90% of susceptibles at each time step, until the number of susceptibles falls below 152. As expected, the maximum expected cost (2297 cost units) is realized through inaction (*top* row and left column policies are never implemented)
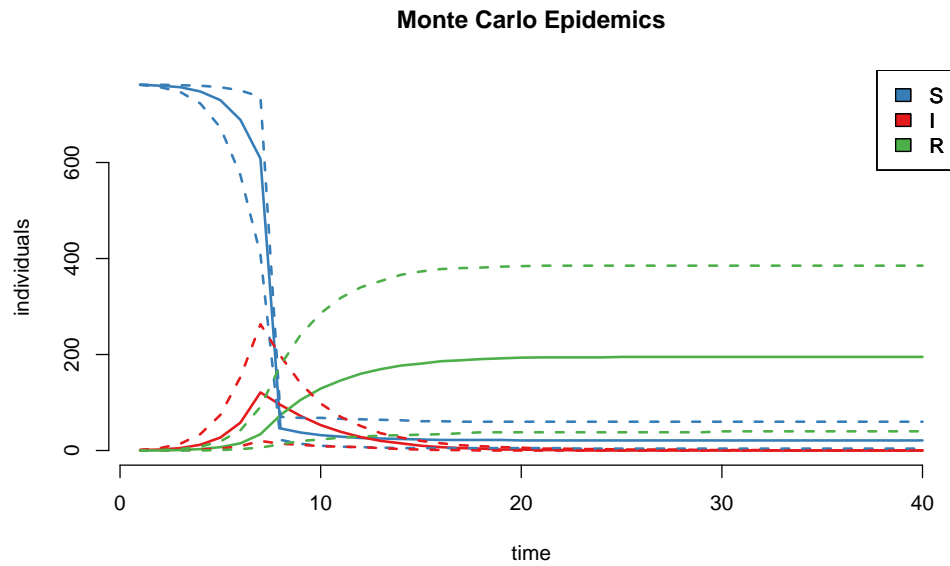
```
R> getvac(opt.MCepi)
```

```
  q0.025   mean median q0.975
1    368 538.38    548    664
```

We can compare these results to the case without vaccination (Figure 3, *top*), and see that the optimal vaccination strategy effectively suppresses the spread of the infection. The costs also spike around time 7 as the vaccination policy is implemented. These costs then stabilize at a lower level than that observed under the no–vaccination strategy in Figure 1 (*bottom*). We can easily extract information on the distribution of final costs of the no–vaccination and optimal (fixed) vaccination policies for comparison as follows.

```
R> T <- length(opt.MCepi$Median$C)
R> optC <- getcost(opt.MCepi)
```

```
R> plot(opt.MCepi)
```

Monte Carlo Epidemics


```
R> plot(opt.MCepi, type = "costs")
```
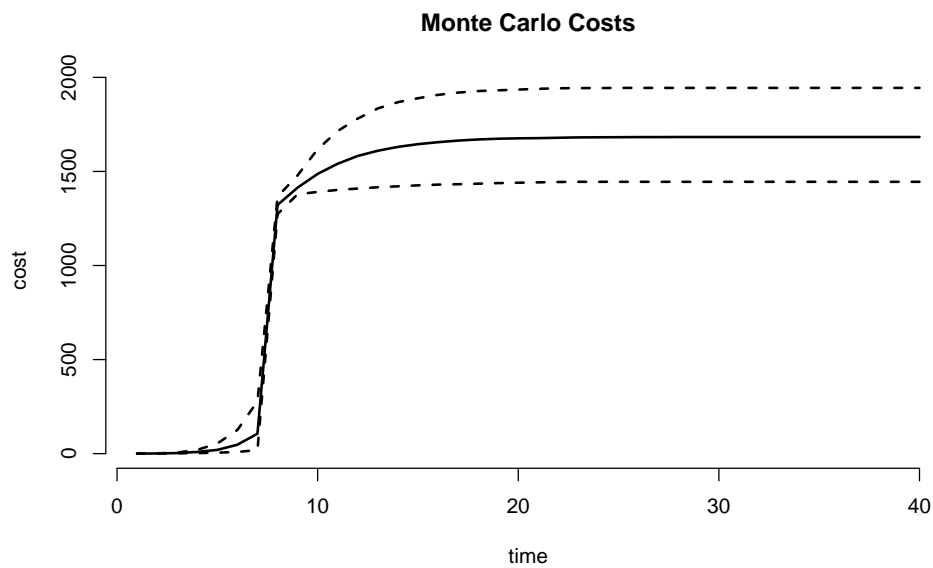
Monte Carlo Costs

Figure 3: Monte Carlo simulated trajectories (*top*) for the numbers of susceptible, infected, and recovered individuals and cost(s) (*bottom*) trajectories under optimal (fixed) vaccination strategy.

```
R> initC <- getcost(init.MCepi)
R> data.frame(rbind(initC, optC), row.names = c("init",
```

```
+       "opt"))
```

```
     q0.025   mean median q0.975
init 2142.8 2297.1   2299 2437.0
opt  1444.9 1684.1   1683 1944.1
```

The optimal (fixed/static) strategy gives a (mean) savings of approximately 613.004 units, or 27% compared to the cost without an intervention. The same information is available through the generic `print` and `summary` commands, e.g.

```
R> opt.MCepi
```

```
Call:
MCepi(init = init, params = tp, vac = vac.opt, costs = costs)
```

```
Distribution of vaccinations administered:
    q0.025   mean median q0.975
vac    368 538.38    548    664
```

```
Distribution of final costs:
     q0.025   mean median q0.975
cost 1444.9 1684.1   1683 1944.1
```

Now we consider adaptive management strategies using the `manage` function. In this case we assume that we do not have perfect information on the underlying epidemic. As such, we want to simultaneously estimate the epidemic model parameters as well as find an optimal management strategy.

In order to proceed, we must define the `epistep` function, which will provide the "observed" SIR count data upon which our inferences will be based. This function can be regarded as the "true" disease process that is being approximated by the SIR model we have described. The default `epistep` function provides simulations from the same disease model (Eqns. 5–7) that is being used for assumed for inference, however the user can specify custom `epistep` functions. By customizing this function, a user can provide custom data to **amei**, or explore the effects of model misspecification, as we shall describe in Section 4.

We also need to start with an initial guess, i.e. priors, for the epidemic parameters that we want to estimate. The default option is to do this by choosing appropriate hyperparameters for the priors explained earlier. The defaults used by `manage` are those given in Section 2.2.

Here, we run the `manage` function with default values for hyperparameters and `epistep` to adaptively design a vaccination strategy to manage the epidemic. At each time step in the evolution of the epidemic the `manage` function uses MCMC to sample from the posterior distribution of the parameters (b, k, ν, μ) given the available history of the epidemic and any already–implemented intervention. Then, a thinned subset of these samples are used propagate uncertainty in the parameter estimates through to the costs of the vaccination strategies. These costs are obtained by performing Monte Carlo forward simulations of the epidemic from the current time point into the future with those parameters. As explained in Section 2.3 we

choose to implement the strategy that most frequently minimizes the cost. After the intervention is implemented, the state in the next time step is determined by `epistep`, and the process is repeated.[4]

```
R> out.man <- manage(init, epistep, vacgrid, costs)
```

To explore the results of the simulation we again plot the evolution of the epidemic (Figure 4, *top*) as well as the cost trajectory under the optimal vaccination (Figure 4, *bottom*), both for a single run. We can now compare the case with adaptive management to the case without vaccination (Figure 1, *bottom*), as well as the case of optimal vaccination with perfect information (Figure 3, *bottom*). As before, we can extract information on the final cost of the epidemic as follows.

```
R> getcost(out.man)
```

```
[1] 1744
```

Since the `manage` function is not performing a Monte Carlo experiment, the number returned is a scalar. Notice that the adaptive strategy is comparable to the best fixed strategy obtained when the true parameterization is known. Later, we shall perform a Monte Carlo version of the adaptive management strategy to make a more meaningful comparison.

We can also see the final distribution of the SIR model parameters contained in `out.man$samps`. This is shown in Figure 5. A summary is provided by the generic `print` and `summary` commands:

```
R> out.man
```

```
Call:
manage(init = init, epistep = epistep, vacgrid = vacgrid, costs = costs)

Distribution of SIR model parameters:
      b                 k                 nu
 Min.   :0.00189   Min.   :0.0842   Min.   :0.271
 1st Qu.:0.00233   1st Qu.:0.2347   1st Qu.:0.330
 Median :0.00250   Median :0.3481   Median :0.346
 Mean   :0.00252   Mean   :0.4461   Mean   :0.348
 3rd Qu.:0.00268   3rd Qu.:0.5710   3rd Qu.:0.366
 Max.   :0.00379   Max.   :2.0773   Max.   :0.443
      mu
 Min.   :1.21e-06
 1st Qu.:5.30e-04
 Median :1.39e-03
 Mean   :2.07e-03
 3rd Qu.:3.09e-03
```
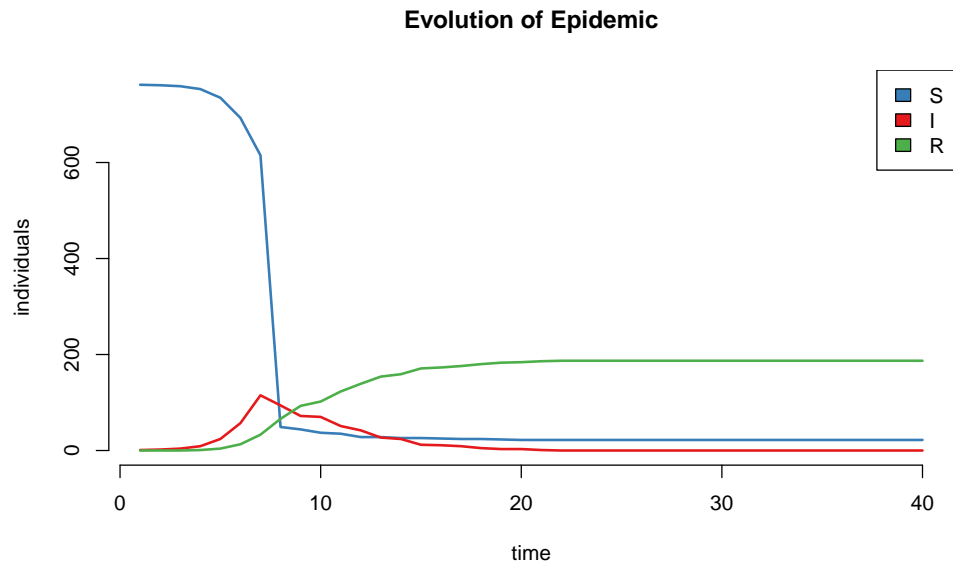
---

[4]The function `manage` has default time horizons specified for both the last time an intervention can be implemented and the time of the final observation of the epidemic. These can be varied with the arguments `Tstop` and `T`, respectively. By default we specify `T=40` and `Tstop=T`.

```
R> plot(out.man)
```

Evolution of Epidemic


```
R> plot(out.man, type = "cost")
```
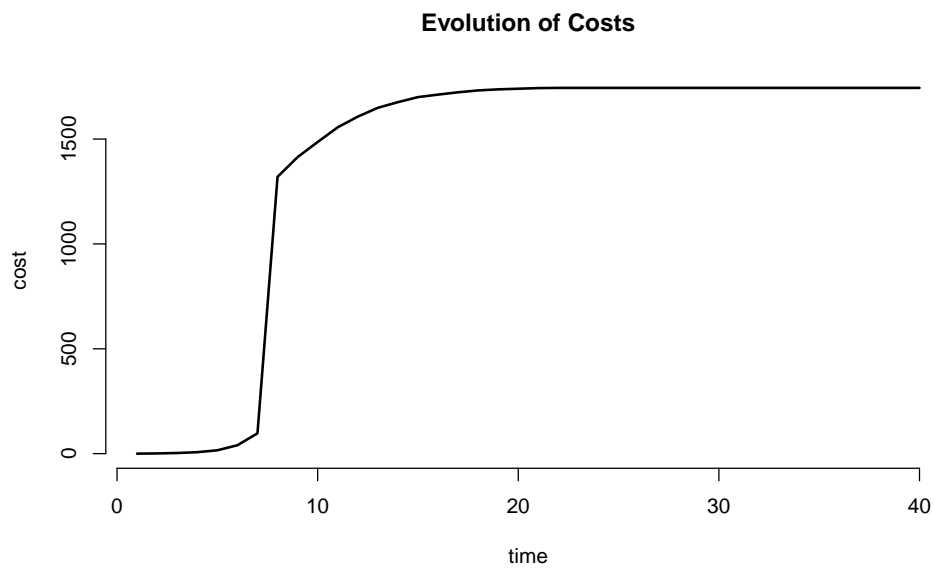
Evolution of Costs

Figure 4: Trajectory (*top*) in terms of the numbers of susceptible, infected, recovered, and vaccinated individuals, and the corresponding cost (*bottom*) of the epidemic under adaptive management.

```
 Max.   :1.16e-02
```

```
R> true <- as.list(formals(epistep)$true)
R> plot(out.man, type = "params", tp = tp)
```
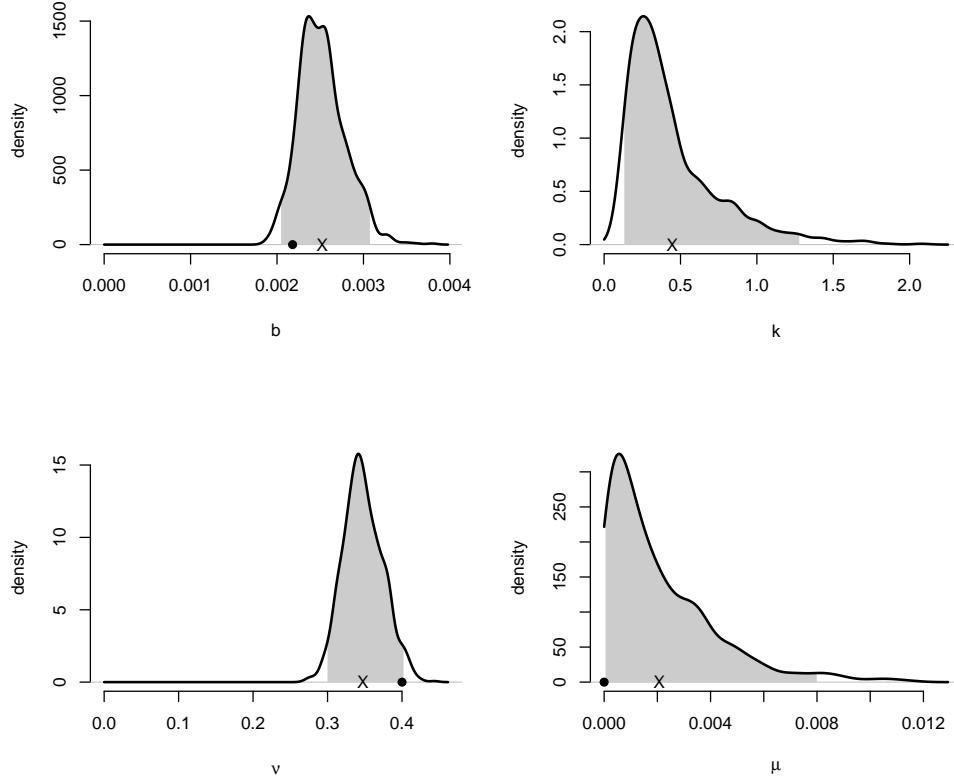

Figure 5: Final posterior distributions of estimated parameters (clockwise from the left: transmission rate, overdispersion parameter, mortality rate, recovery rate). "True" parameter values are indicated by a dot; mean posterior values are indicated by an 'x'; and the central 95% region of the distribution is shaded.

```
Vaccinations administered and cost:
  vac cost
1 554 1744
```

We also note that it is possible to utilize the `manage` function in a piecemeal way, or in "epochs". This is accomplished by passing the output of the `manage` function, an `"epiman"`-class object, as the first (`init`) argument. The management of the epidemic will re-start where it left off. It will stop at the newly specified (later) time horizon `Tstop` specified by that argument.

Next, we illustrate a Monte Carlo experiment where epidemics are initialized and proceed randomly through the adaptive management strategy illustrated above so that we can see the average behavior, costs, and associated variability. The function `MCmanage` facilitates this experiment, and it essentially calls the `manage` function many times (which can be controlled by the `MCreps` argument). In order to provide a (relatively) quick demonstration we have set a

low default of `MCreps = 30` and have used low defaults for the other Monte Carlo parameters to `management`.

```
R> out.MCmanage <- MCmanage(init, epistep, vacgrid, costs)
```

To reproduce the results in Merl *et al.* (2009), use

```
MCvits = 100, MCMCpits = 10000, vacsamps = 100, MCreps = 100
```

and otherwise use the defaults. The object that is returned is of class `"MCepi"` with fields similar to those that are output from the `MCepi` function which implements a static (fixed) vaccination strategy. Thus, the same generic plot commands can be used. Figure 6 shows plots summarizing the distribution of epidemic trajectories (*top*) and costs (*bottom*) under the adaptive management. Distributional information on number of vaccine units dispensed can be extracted as follows.

```
R> getvac(out.MCmanage)


   q0.025    mean median q0.975
1 454.35 554.77    551 663.12
```

Figure 7 shows the distribution of fractions of individuals vaccinated and the stopping level for the epidemic trajectories under adaptive management in the Monte Carlo experiment. We can compare the costs (and quantile bounds) to that of the best fixed vaccination strategy calculated assuming that the true parameterization is known.
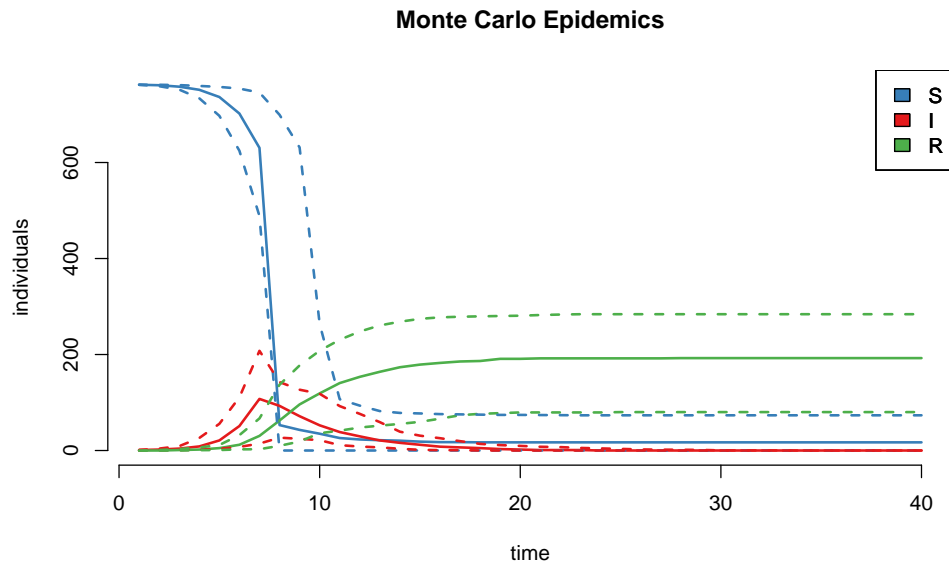
```
R> cinit <- getcost(init.MCepi)
R> copt <- getcost(opt.MCepi)
R> cman <- getcost(out.MCmanage)
R> data.frame(rbind(cinit, copt, cman), row.names = c("init",
+     "opt", "man"))


      q0.025   mean median q0.975
init 2142.8 2297.1 2299.0 2437.0
opt  1444.9 1684.1 1683.0 1944.1
man  1494.2 1690.2 1687.5 1855.5
```

The cost of the non-intervention strategy is shown again for calibration purposes. As before, the generic `print` and `summary` commands can be used to obtain the same information. Notice that, in this short Monte Carlo experiment, the adaptive management scheme approximates the best static vaccination scheme obtained then the true parameterization is known in advance.

It is interesting to compare to what the cost of managing the epidemic would have been *without* estimating the parameters as the epidemic progressed, but rather by guessing what the appropriate parameters might be. Suppose our best guess at the parameters underestimated the true transmission probability b and overestimated the true recovery probability ν.

```
R> plot(out.MCmanage)
```

Monte Carlo Epidemics


```
R> plot(out.MCmanage, type = "costs")
```
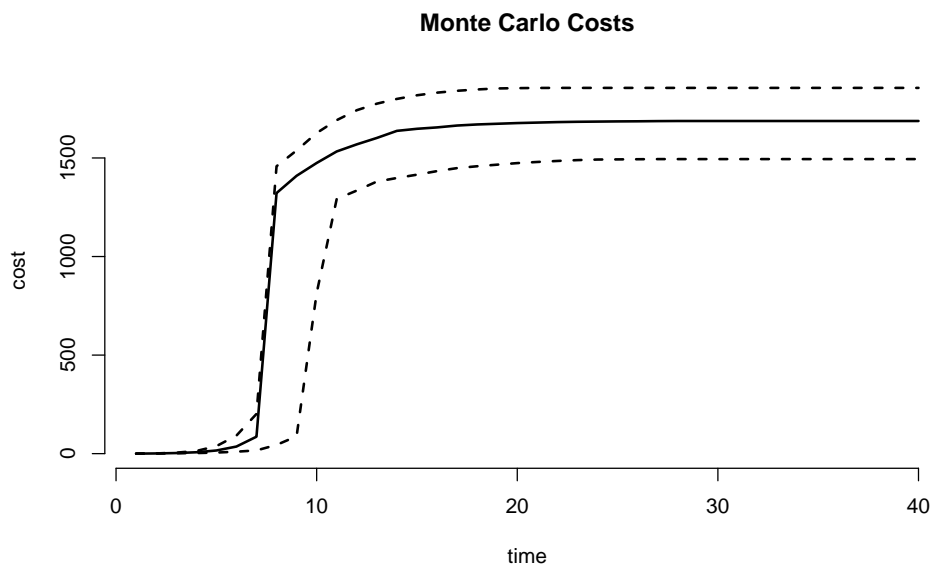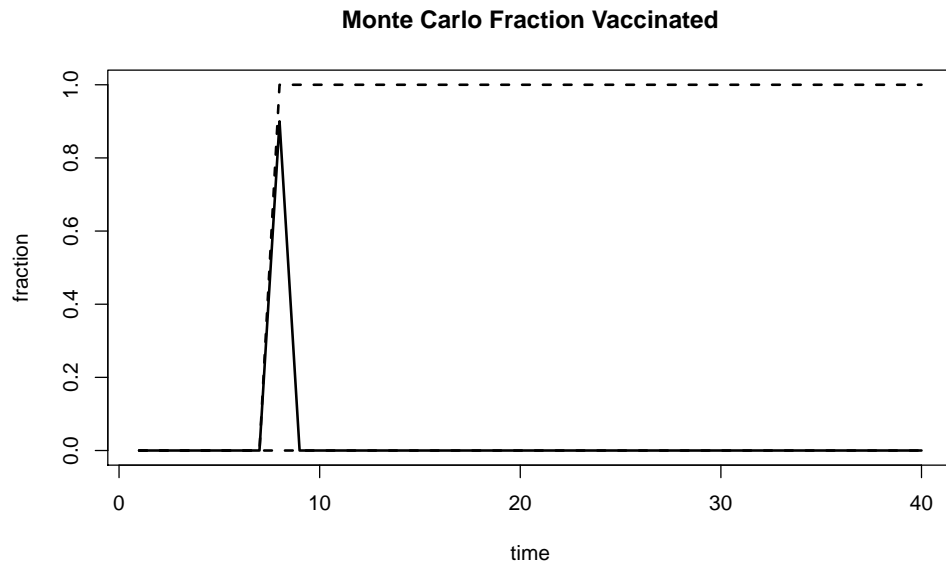
Monte Carlo Costs

Figure 6: Monte Carlo simulated epidemic trajectories (*top*) for the numbers of suscepti-
ble, infected, and vaccinated individuals, and the associated cost(s) (*bottom*) under adaptive
management

```
R> bad <- list(b = 0.001, k = 10, nu = 0.9, mu = 0)
```

The optimal (static) policy under this parameterization can be constructed, as demonstrated

```
R> plot(out.MCmanage, type = "fracs")
```

Monte Carlo Fraction Vaccinated


```
R> plot(out.MCmanage, type = "stops")
```

Monte Carlo Stopping Threshold

Figure 7: Fractions vaccinated (*top*) and stopping levels (*bottom*) for the simulated epidemic trajectories shown in Figure 6 that were under adaptive management. In both cases, the $(2.5, 50, 97.5)$-% quantiles are shown.

above.

```
R> costs.bad <- optvac(init, bad, vacgrid, costs)
```

```
R> pol.bad <- getpolicy(costs.bad)
R> pol.bad
```

```
     row col frac stop cost
best   2  10  0.1  677  376
```

Then, we can calculate the distribution of costs of managing the true epidemic with a policy developed under our best guess of the parameterization.

```
R> bad.MCepi <- MCepi(init, tp, pol.bad[3:4], costs)
R> cbad <- getcost(bad.MCepi)
R> cbad
```

```
  q0.025   mean median q0.975
1   2126 2278.6 2280.5   2436
```

Comparing these costs with the ones obtained above, under adaptive management, we can see that a poor guess can lead to a significantly worse strategy—nearly 57% larger on average. Clearly, good quality (online) estimates of the SIR model are crucial to ensuring a cost–effective approach to the management of an epidemic.

4. Alternative Transmission Model

We would like to see how well the fixed and adaptive strategies can do when faced not only with parameter mis-specification, but when there may be some component of the underlying transmission model that is not accounted for in the implemented SIR model upon which its vaccination strategies are based. Towards this end, we chose a fairly simple extension of the SIR model where infection does not pass directly from individual to individual. Instead, they become infected by encountering a reservoir of the infectious agent (for instance bacteria or fungi in water or soil) contributing infected individuals. The continuous time dynamics are:

$$
\begin{aligned}
\frac{dS}{dt} &= -aS\left(\frac{C}{C+C_0}\right) \\
\frac{dI}{dt} &= aS\left(\frac{C}{C+C_0}\right) - (\nu + \mu)I \\
\frac{dR}{dt} &= \nu I \\
\frac{dD}{dt} &= \mu I \\
\frac{dC}{dt} &= \rho I - mC
\end{aligned}
\tag{14}
$$

where the concentration of the infective agent in the reservoir is given by C; the transmission rate is modeled by a saturating function of C, so that as $C \to \infty$ the transmission rate approaches the constant a at a rate determined by C_0; infective agents die or are removed from the reservoir at rate mC; and the *per capita* rate at which new infected agents are added to the reservoir is ρ.

This system can be discretized in a similar manner to the system discussed earlier, so that the single step transmission dynamics are given by:

$$\tilde{I}|s, C \sim \mathrm{Bin}(s, p_i(a, C, C_0)), \tag{15}$$

where

$$p_i(a, C, C_0) = 1 - \exp\left(\frac{-aC}{C + C_0}\right) \tag{16}$$

and $\mathrm{Bin}(n, \pi)$ is the binomial distribution. The dynamics for $\tilde{R}$ and $\tilde{D}$ are exactly as before (Eqs. (6–7)). The single step dynamics for the reservoir are expected to be more smooth than the epidemic in the population at large. For instance, for bacteria the concentration could be in the thousands or millions of individuals. In this case the evolution of the reservoir is approximately

$$C(t+1) - C(t) = [\rho I(t)] - d_c. \tag{17}$$

Here d_c is the number (per unit reservoir) of infectious agents (stochastically) removed from the reservoir in a unit time, which has distribution

$$d_c \sim \mathrm{Bin}(C, p_{d_c}), \tag{18}$$

where $p_{d_z} = 1 - e^{-m}$.

In R, we may implement the above transition model by encoding it in an alternative `epistep` function for use with the **amei** package methods as follows.

```
R> alt.epistep <-
+ function(SIR, last=list(rem=0, rec=0, infect=0, dead=0, Z=0),
+          tp=list(a = 0.05, mu = 0.05, nu = 0.1, m = 0.4,
+          rho = 200, C = 500))
+ {
+   ## calculate the infection probability based on the
+   ## reservoir, and randomly infect susceptibles
+   Z <- last$Z
+   fz <- Z/(Z+tp$C)
+   pi <- 1 - exp(-tp$a * fz)
+   infect <- rbinom(1, SIR$S, pi)
+
+   ## update recovereds and deaths
+   pr <- 1 - exp(-tp$nu)
+   rec <- rbinom(1,SIR$I,pr)
+   pd <- 1 - exp(-tp$mu)
+   dead <- rbinom(1, SIR$I-rec, pd)
+
+   ## reservoir dynamics
+   pz <- 1 - exp(-tp$m)
+   dz <- rbinom(1, Z, pz)
+   bz <- round(SIR$I*tp$rho)
```

```
+   Z <- Z - dz + bz
+
+   ## the returned list is passed in as "last" in a
+   ## subsequent call to this "epistep" function
+   return(list(rem=(rec+dead), rec=rec, infect=infect,
+               dead=dead, Z=Z))
+ }
```

Here we first use the `manage` function with a `NULL` vaccination strategy (and also, optionally, a `NULL` cost structure), in order to see how the behavior of this system compares to the default `epistep` model implemented within **amei**. We can then also look at the estimated "effective" SIR parameters.

This system can resemble the standard SIR model, especially when m is large, so infective agents do not remain in the reservoir for long.

```
R> init1 <- list(S0 = 150, I0 = 1, R0 = 0, D0 = 0)
R> tp <- list(a = 0.1, mu = 0, nu = 0.3, m = 50, rho = 500,
+      C = 500)
R> alt.epistep1 <- alt.epistep
R> formals(alt.epistep1)$tp <- tp
R> out.alt <- manage(init1, alt.epistep1, NULL, NULL, T = 80)
```

The *top* of Figure 8 shows the resulting dynamics under the default parameterization offered by the formals of the `alt.epistep` function. To illustrate how this new system can (significantly) differ from the SIR dynamics consider the case of very small m. Small m enables new infections to occur even if there had been no infected individuals in a previous time step since the infective agent may persist in the reservoir for a long time without infected individuals being present.

```
R> tp <- list(a = 0.1, mu = 0, nu = 0.3, m = 0.001, rho = 500,
+      C = 500)
R> alt.epistep2 <- alt.epistep
R> formals(alt.epistep2)$tp <- tp
R> out.alt2 <- manage(init1, alt.epistep2, NULL, NULL, T = 80)
```
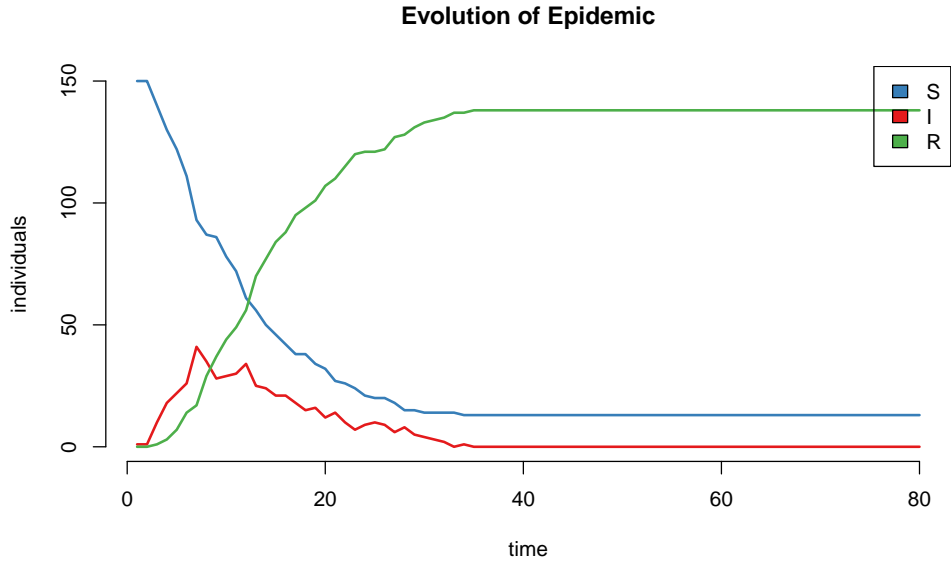
The effect of these new dynamics may be readily seen in the *bottom* panel of Figure 8 as the epidemic progresses. In this second case (with small m), it is also very unlikely that any susceptible individuals will be left at the end of an epidemic. This contrasts with both the stochastic SIR model and these alternative dynamics with m large, as it is possible that, due to stochastic effects, the infection will die out before all susceptibles have been exposed.

It is interesting to examine the posterior distributions of the parameters for these two cases (Figure 9) if we assume that these underlying alternative dynamics are well approximated by the simpler SIR model. As one might expect, the estimates of the recovery and death rates in both cases are quite similar, especially since these dynamics are the same as those implemented within **amei**. However, the estimates of b and k are quite different in the two cases. The 95% credible intervals hardly overlap, indicating that the two cases result in dynamics that are quantitatively different.

```
R> plot(out.alt, showv = FALSE)
```

Evolution of Epidemic


```
R> plot(out.alt2, showv = FALSE)
```
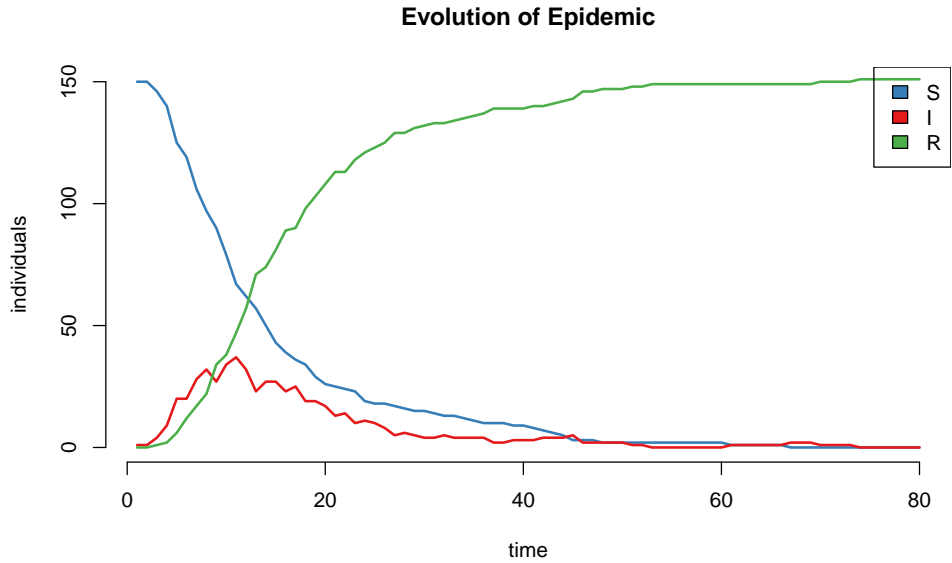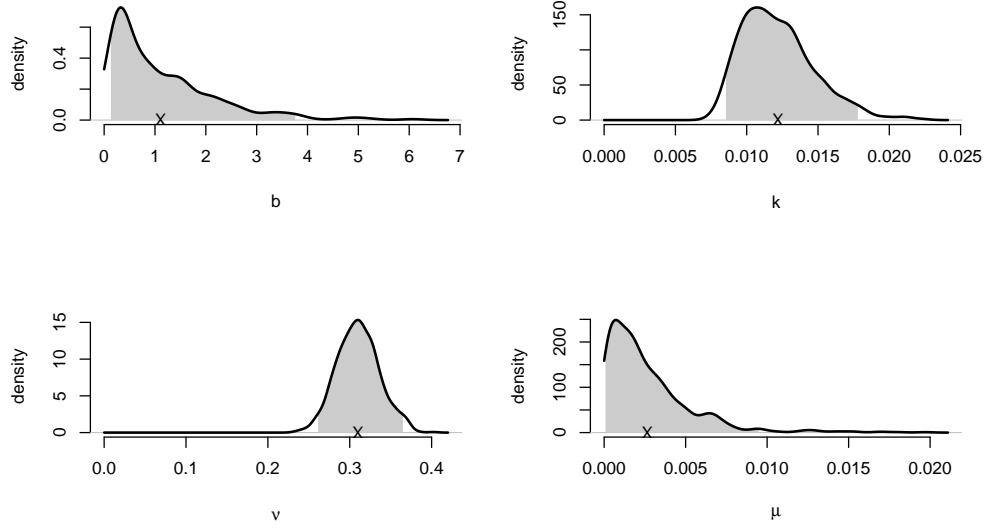
Evolution of Epidemic

Figure 8: Trajectories in terms of the numbers of susceptibles, infecteds, recovereds, and dead individuals under the alternative model in Eq. (14) with parameters $\{(a = 0.1, \mu = 0.0, \nu = 0.3, \rho = 500, C_0 = 500\}$ and with (*top*) $m = 50$ (*bottom*) $m = 0.001$.

We now move on to our task of examining how well the fixed and adaptive strategies can do when faced with an epidemic which is evolving according to a transmission function outside of the class of SIR models (1–3) used to calculate the (optimal) vaccination strategy. The first step here is to let the fixed strategy "cheat". That is, we allow the fixed strategy to see a full epidemic spread according to the true model and parameterization without intervention, and estimate the best SIR model approximation. To do this we use the `manage` function with

```
R> plot(out.alt, type = "params", showd = TRUE)
```
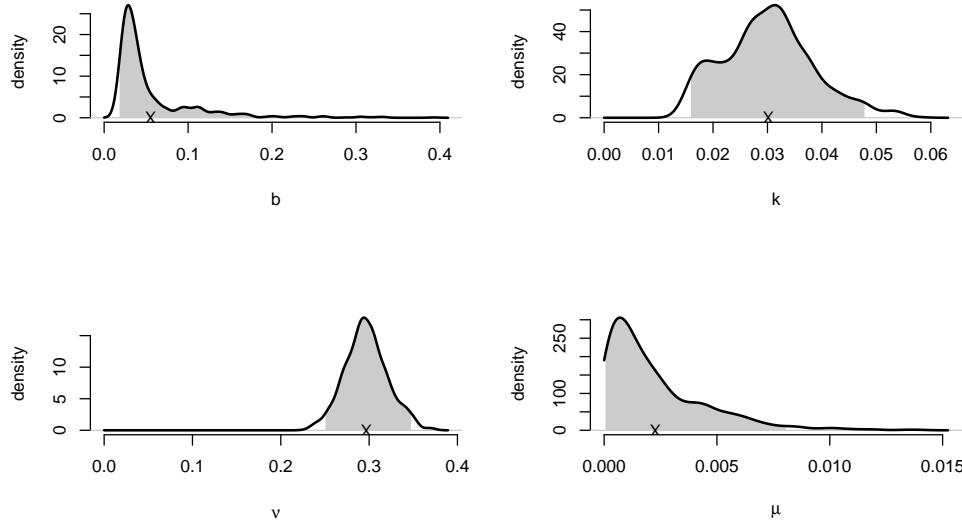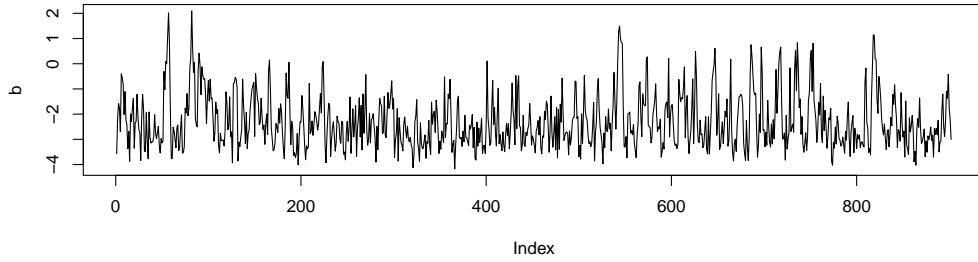


```
R> plot(out.alt2, type = "params", showd = TRUE)
```


Figure 9: Posterior parameter distributions for b, k, μ, and ν when the simulated source epidemic follows the alternative transmission model with parameters $\{(a = 0.1, \mu = 0.0, \nu = 0.3, \rho = 500, C_0 = 500\}$ and with (*top*) $m = 50$ (*bottom*) $m = 0.001$.
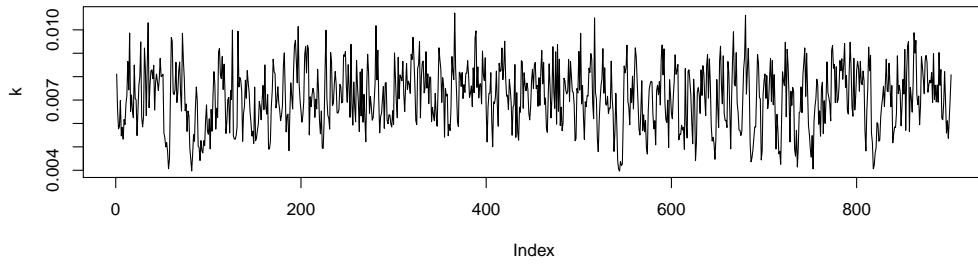
a NULL vaccination strategy and a NULL cost structure, and default "true" parameters.

```
R> init <- list(S0 = 600, I0 = 1, R0 = 0, D0 = 0)
R> time <- 80
R> posterior <- manage(init, alt.epistep, NULL, NULL, T = time,
+      bkrate = 100)
```
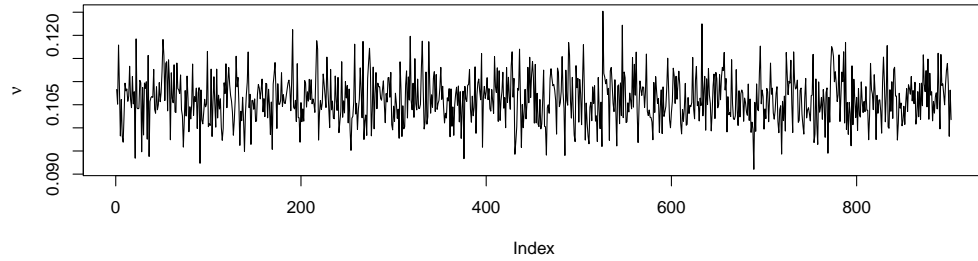
```
R> plot(log(posterior$samp$b), type = "l", main = "", ylab = expression(b))
```



```
R> plot(posterior$samp$k, type = "l", main = "", ylab = expression(k))
```



```
R> plot(posterior$samp$nu, type = "l", main = "", ylab = expression(nu))
```



```
R> plot(posterior$samp$mu, type = "l", main = "", ylab = expression(mu))
```
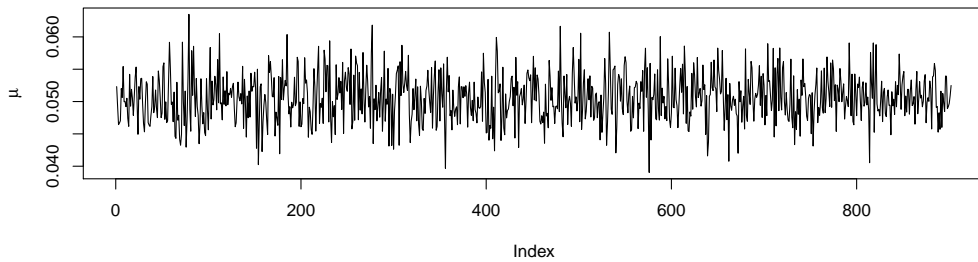

Figure 10: Traces of samples from the posterior distribution of the parameters for the (misspecified) SIR model.

Figure 10 shows trace plots of the samples obtained. From the output we can extract the mean parameterization obtained at the final time point of the epidemic.

```
R> mean.params <- as.list(apply(posterior$samp, 2, mean))
```

Based on these parameters, and thus assuming an SIR model, we can calculate the optimal static vaccination policy.

```
R> costs <- list(vac = 2.5, death = 4, infect = 1)
R> vacgrid <- list(fracs = seq(0, 1, 0.1), stops = seq(2,
+     init$S0 - 50, 50))
R> alt.optvac <- optvac(init, mean.params, vacgrid, costs,
+     T = time)
R> alt.best <- getpolicy(alt.optvac)

R> plot(alt.optvac)
```
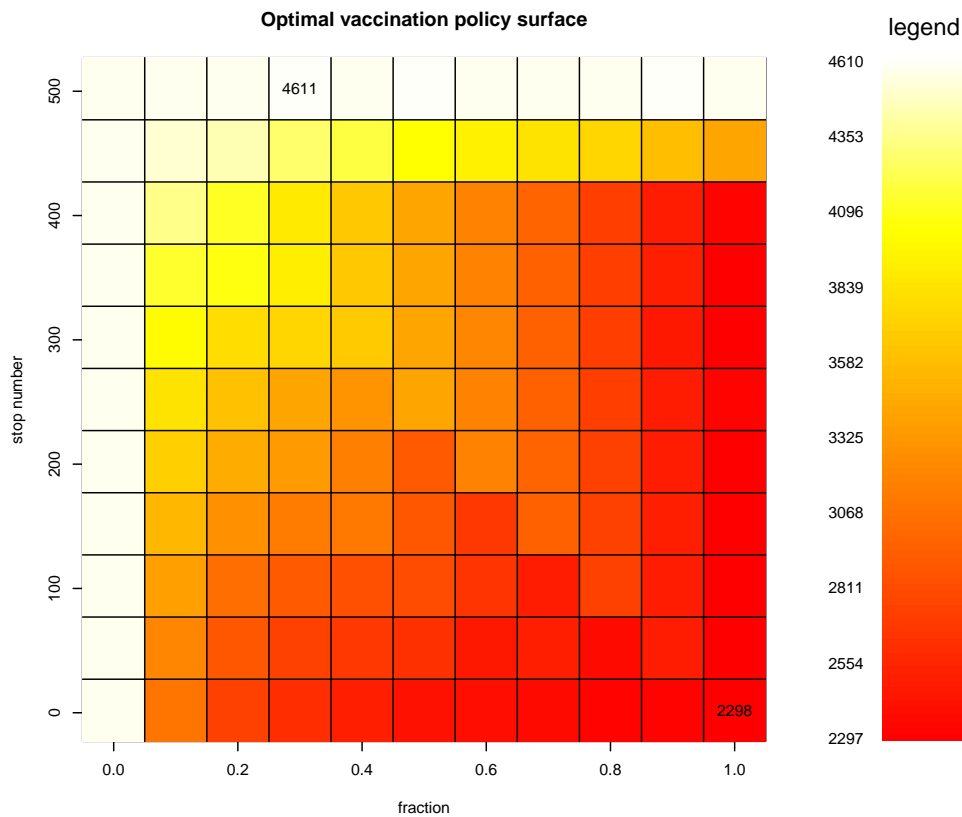

Figure 11: The cost surface calculated for fixed strategies using parameters estimated from a single run of the alternative epidemic model.

The cost surface is shown in Figure 11.

Then, based on the extracted policy, we can use `MCepi` to simulate the true evolution of the epidemic (via `alt.epistep`) many times in order to build up an understanding of the distribution of costs of the optimal static policy calculated under the simplified SIR parametrization.

```
R> alt.vac.opt <- alt.best[3:4]
R> alt.MCepi <- MCepi(init, alt.epistep, alt.vac.opt, costs,
+     T = time)

R> getcost(alt.MCepi)
```

```
  q0.025    mean median q0.975
1   1854 2027.6    2026 2216.1
```

Now for the comparison. We provide the `alt.epistep` function to `MCmanage` in order to build up an understanding of how the optimal adaptive strategy fares under a restricted SIR model. Here we use the default, non-informative priors.

```
R> alt.MCmanage <- MCmanage(init, alt.epistep, vacgrid,
+     costs, T = time)

R> getcost(alt.MCmanage)

  q0.025    mean median q0.975
1 1829.1 2026.9 2039.2 2185.4
```

An alternative (and possibly fairer) approach would be to allow the adaptive strategy to cheat as well by choosing the prior to be tightly concentrated around around the `mean.params` estimated above. We can see by looking at the full trajectories in Figures 12 and 13 that the strategies implemented by both the fixed and adaptive methods are very similar. The adaptive strategy can do better than the fixed strategy, both in extreme and average cases. The adaptive framework has the added benefit that prior information on similar epidemics may be used, which typically results in a narrower spread of (often lower) costs, leading to even further improvements over fixed strategies—even ones which can see into the future.

We can also look back to Figure 11 to see the kind of improvement that the optimal strategy can have over the case when no one is (or very few individuals are) vaccinated. We can extract the expected cost under this worst fixed strategy and compare it to the best fixed strategy directly.

```
R> alt.worst <- getpolicy(alt.optvac, which = "worst")
R> rbind(alt.best, alt.worst)

      row col frac stop cost
best   11    1  1.0    2 2298
worst   4   11  0.3  502 4611
```
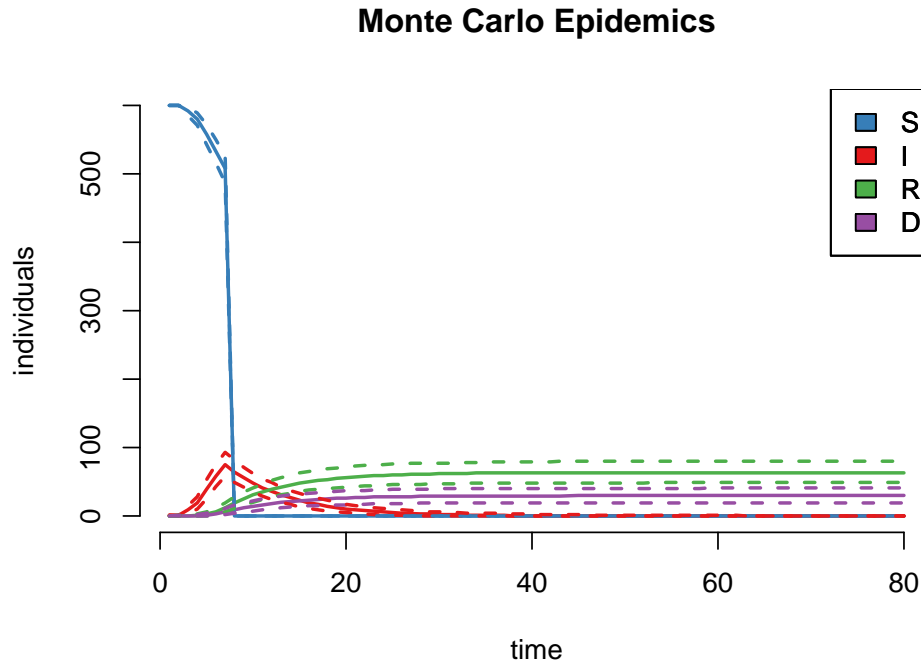
Here it is easy to see how much better using an optimal fixed strategy with good parameter settings is compared to not vaccinating. Since the adaptive strategy has comparable (or lower) costs to the fixed strategy over many trials of the epidemic, we can conclude that the adaptive strategy gives similar reductions in costs over the `NULL` vaccination policy as the fixed policy does, even though the adaptive algorithm must begin planning and implementing the intervention with considerably less information than the fixed strategy it is compared with here.

We have shown here that it is straightforward to build functions with fairly simple extensions to the SIR model to use with the **amei** package. Using one example, we can see that the model implemented within **amei** to plan the interventions is flexible enough to allow effective vaccination strategies, even when this model does not match the one which is known to govern

```
R> plot(alt.MCepi, showd = TRUE)
```

Monte Carlo Epidemics


```
R> plot(alt.MCepi, type = "costs")
```

Monte Carlo Costs

Figure 12: Trajectories in terms of *top* the numbers of susceptibles, infecteds, recovereds, and dead individuals under the alternative model in Eq. (14) and *bottom* costs under the default parameterization under the optimal static vaccination policy.

```
R> plot(alt.MCmanage, showd = TRUE)
```

Monte Carlo Epidemics


```
R> plot(alt.MCmanage, type = "costs")
```
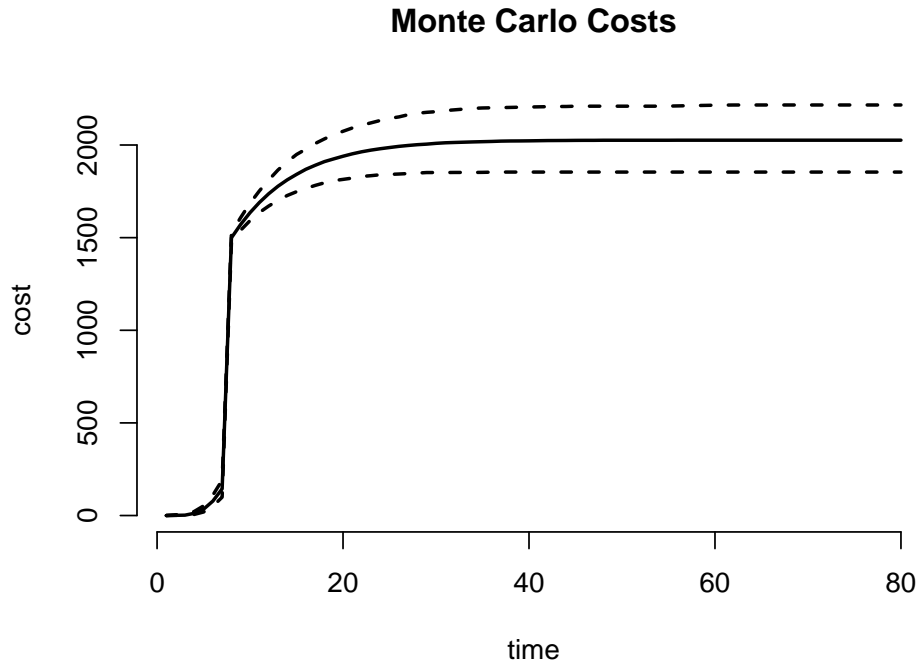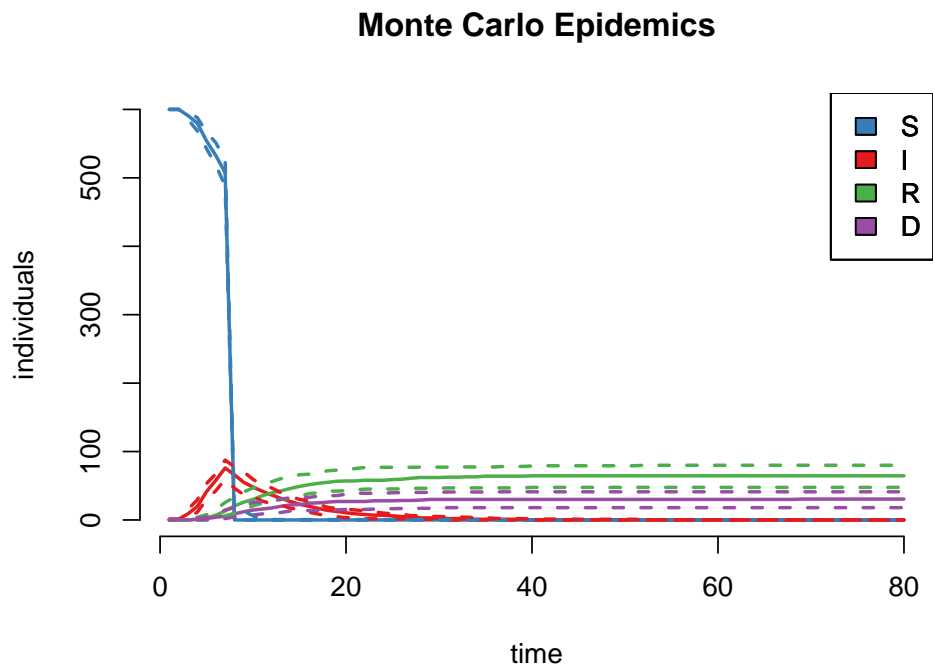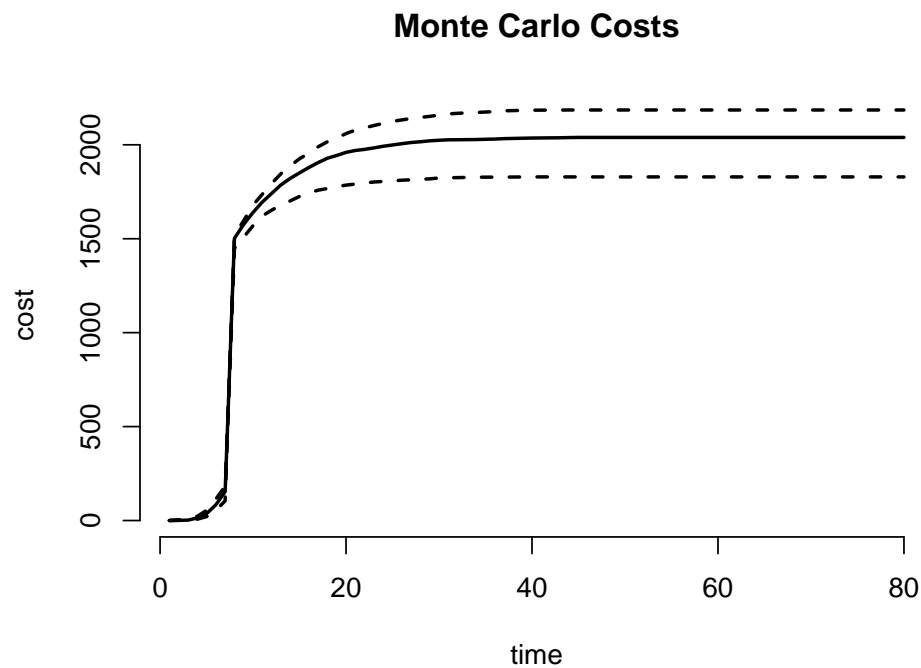
Monte Carlo Costs

Figure 13: Trajectories in terms of *top* the numbers of susceptibles, infecteds, recovereds, and dead individuals under the alternative model in Eq. (14) and *bottom* costs under the default parameterization under the adaptive vaccination policy.

the true underlying epidemic dynamics. We have also shown how one can estimate "effective" SIR parameters from an alternative dynamic epidemic model. These estimates can be used to build a fixed policy, or the posterior samples could even be used as parameters to enhance the adaptive strategies.

5. Discussion and related work

Our **amei** package for R implements a statistical framework that enables concurrent estimation of epidemic parameters and optimal intervention strategies. In particular, it allows parameter uncertainty to be taken into consideration when planning an intervention.

In the current implementation, we look for adaptive strategies of only one type—vaccination of proportions of the population until susceptibles fall below some threshold. However, if we were to instead allow the fraction of the population targeted for vaccination to be a function of future disease states we could regard Eq. (13) as a stochastic iteration equation and use stochastic dynamic programming (Clark and Mangel 2000) to calculate the optimal intervention associated with a set of parameter values. Such an approach may be useful for situations in which knowledge of the disease state is available, but for whatever reason sequential inference is not possible. In the situation considered here, in which the static strategy is sequentially updated based on the current disease state and parameter estimates, the adaptive strategy that emerges is flexible in that it consists of a state–dependent sequence of target fractions, but does not involve the additional computational burden associated with stochastic dynamic programming.

As in many decision problems, it can be difficult to specify cost functions that truly represent all aspects of the actions. Although **amei** is based on a very simple cost evaluation scheme, it is straightforward to extend our approach to more complex cost functions, including state-dependent costs. We refer the reader to Merl *et al.* (2009) for more detailed discussion of more complex cost assignments.

If there is not much data observed during an epidemic (for instance if the epidemic is occurring in a small population) then estimating disease parameters within any framework could, of course, be difficult. One could use data about previous epidemics to estimate parameters, for instance by maximum likelihood methods, to plan an intervention. However, as we showed in Section 3, should these estimates be poor then the cost of intervening in the epidemic can be greatly increased. By instead using previous epidemic data to inform the priors within a Bayesian framework, such as the one we describe here, we can take into account other research and observations without risking the high costs inherent in using incorrect parameter estimates for planning an intervention.

Recently other authors (e.g. Elderd *et al.* 2006) have also proposed using Bayesian methods to estimate disease parameters and propagate uncertainty in parameter estimates through to an optimal of vaccination strategy. For instance, Elderd *et al.* (2006), considered a Susceptible–Exposed–Infectious–Recovered (SEIR) model for a small pox epidemic with mass action infection dynamics, and either mass or trace vaccination. However, in their approach the likelihood requires numerical solutions of the system of DEs (which can be computationally intensive). Ball and Lyne (2002); Ball *et al.* (2007) have also written technical papers on the optimal vaccination strategies in epidemics, but do not consider adapting the policy over time.

Compared to these approaches, our proposed method has several advantages. Our negative

binomial discretization of the SIR model is much less computationally intensive, and could easily be expanded to include an exposed class. Additionally, this model is comparatively simple since it connects intuitively with a straightforward SIR model, and is thus likely to be much more approachable to policy makers and practitioners outside of statistics. Our approach is also adaptive, allows a much more flexible cost framework, and is generalizable to other types of interventions. Since the intervention strategy implemented by **amei** is iteratively updated upon the arrival of new data, the vaccination schedule is inherently adaptive to the state of the epidemic. This allows significant changes in the vaccination strategy mid-intervention. This may be helpful, for example, in a scenario where an intervention may be discontinued during a lull, a subsequent refinement of parameter estimates or a surge of new infections may dictate that the vaccination campaign should be re-initiated.

Not only can more complicated infection dynamics be used for simulating the dynamics of a disease, but in addition the method implemented in **amei** can be modified to include more complicated disease dynamics such as latent states or vector-communicated diseases, as well as more complicated intervention strategies that allow combinations of vaccination, quarantine, and culling. Also note the possibility of calculating policies based on a minimization of some quantile of the realized cost rather than the mean cost. This would lead to a minimization of worst-case-scenarios, which may be useful in practice.

There are other R packages that involve the simulation of epidemics, and inference for models like SIR. As far as we can tell, ours is the only one which considers (on–line) adaptive management of interventions. However, two packages that take approaches similar to ours for inferring the parameters governing an epidemic are worth mentioning here. **stochasticGEM** (Zwane 2007) provides Bayesian inference for partially observed stochastic epidemics. The implementation in this package also allows for estimating parameters governing the infectious and incubation period length. Several variants of the general epidemic model are considered (e.g. Gibson and Renshaw 1998; Höhle *et al.* 2005; O'Neill and Roberts 1999; O'Neill and Becker 2001; O'Neill 2002; Streftaris and Gibson 2004), including the stochastic SIR with Markovian and non–Markovian infectious periods, and SEIR models. As with **amei**, estimation is via MCMC. Höhle, et al., whose methods are implemented in the **stochasticGEM** package, have themselves released an R package called **RLadyBug** (Höhle and Feldmann 2008), which requires JAVA. In **RLadyBug** maximimum likelihood and Bayesian inference can be performed to estimate the parameters and provide confidence/credible intervals and the ability to test hypotheses. The package can also estimate parameters for epidemics which include population heterogeneity. The authors of **RLadyBug** provide a nice paper outlining the usage (Höhle and Feldmann 2007) with examples. Both **RLadyBug** and **stochasticGEM** contain interesting data sets on epidemics and visualization tools.

There are several other R packages that allow for the manipulation and analysis of epidemic and disease data. They include a few packages which are predominantly used for teaching purposes, e.g. : **Epi** (Carstensen *et al.* 2008) with methods for multiscale and censored data; **epiR** (Stevenson *et al.* 2008) focusing on veterinary epidemiology, and **epibasix** (Rotondi 2007). Two other packages that include a range of statistical functionality for data manipulation and inference are **epicalc**(Chongsuvivatwong 2008) and **epitools** (Aragon 2008).

A. Implementation details

At a high level, most of the functions and routines in the **amei** package are written in R. However, most of the sub–routines implementing the Monte Carlo evaluation of costs for the various vaccination strategies, which are obtained by repeatedly simulating the epidemic forward in time given the current samples of the parameter estimates, are written in C for speed considerations. There is significant scope for parallelizations of these Monte Carlo routines, since each forward simulation is independent of the next. For this reason future versions of this package may leverage **Pthreads** or **MPI** to obtain significant speedups.

This document has been authored in **Sweave** (Leisch 2002). This means that the code quoted throughout is certified by R, and the `Stangle` command can be used to extract it. The demo available in this package will run the same code via `demo("amei")`.

Acknowledgments

References

Anderson R, May R (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.

Aragon T (2008). ***epitools**: Epidemiology Tools*. R package version 0.5-2, URL http://www.medepi.com.

Ball F, Lyne O (2002). "Optimal Vaccination Policies for Stochastic Epidemics Among a Population of Households." *Mathematical Biosciences*, **177 & 178**, 333–354.

Ball F, O'Neill P, Pike J (2007). "Stochastic Epidemic Models in Structured Populations Featuring Dynamic Vaccination and Isolation." *Journal of Applied Probability*, **44**, 571–585.

Carstensen B, Plummer M, Laara E, Hills M (2008). ***Epi**: A Package for Statistical Analysis in Epidemiology*. R package version 1.0.8, URL http://www.pubhealth.ku.dk/~bxc/Epi/.

Chongsuvivatwong V (2008). ***epicalc**: Epidemiological Calculator*. R package version 2.7.2.3.

Clark CW, Mangel M (2000). *Dynamic State Variable Models in Ecology: Methods and Applications*. Oxford University Press.

Elderd B, Dukic V, Dwyer G (2006). "Uncertainty in Prediction of Disease Spread and Public Health Responses to Bioterrorism and Emerging Disease." *Proceedings of the National Academy of Sciences (PNAS) USA*, **103**, 15693–15697.

Enserink M (2001). "Rapid Response Could Have Curbed Foot-and-Mouth Epidemic." *Science*, **294**, 26–27.

Ferguson N, Keeling M, Edmunds W, Gani R, Grenfell B, Anderson R, Leach S (2003). "Planning for Smallpox Outbreaks." *Nature*, **425**, 681–685.

Gamerman D, Lopes H (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC.

Gibson G, Renshaw E (1998). "Estimating the Parameters in Stochastic Epidemic Models Using Markov Chain Models." *IMA Journal of Mathematics Applied in Medicine and Biology*, **16**, 19–40.

Godfray H, Hassell M (1989). "Discrete and Continuous Insect Populations in Tropical Environments." *Journal of Animal Ecology*, **58**, 153–174.

Halloran M, Longini I, Nizam A, Yang Y (2002). "Containing Bioterrorist Smallpox." *Science*, **298**, 1428–1432.

Hetchote HW (2000). "The Mathematics of Infectious Diseases." *SIAM Review*, **42**, 599–653.

Höhle M, Feldmann U (2007). "**RLadyBug**–An R Package for Stochastic Epidemic Models." *Computational Statistics and Data Analysis*, **52**(2), 680–686.

Höhle M, Feldmann U (2008). **RLadyBug**: *Analysis of Infectious Diseases Using Stochastic Epidemic Models*. R package version 0.4-3, URL http://www.stat.uni-muenchen.de/~hoehle/software/RLadyBug.

Höhle M, Jørgensen E, O'Neill P (2005). "Inference in Disease Transmission Experiments by Using Stochastic Epidemic Models." *Journal of the Royal Statistical Society Series C*, **54**, 349–366.

Leisch F (2002). "Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis." In W Hardle, B R onz (eds.), *Compstat 2002: Proceedings in Computational Statistics*, pp. 575–580. Physika Verlag.

Lloyd-Smith J, Galvani A, Getz W (2003). "Curtailling Transmission of Severe Acute Respiratory Syndrome Within a Community and its Hospital." *Proceedings of the Royal Society of London*, **270**, 1979–1989.

Mangel M (2006). *The Theoretical Biologist's Toolbox*. Cambridge University Press.

McCallum H, Barlow N, Hone J (2001). "How Should Pathogen Transmission be Modelled." *Trends in Ecology and Evolution*, **16**, 295–300.

Merl D, Johnson L, Gramacy R, Mangel M (2009). "A Statistical Framework for the Adaptive Management of Epidemiological Interventions." *PLOS*, **4**(6), e5807. doi:10.1371/journal.pone.0005807.

Merl D, Johnson LR, Gramacy RB, Mangel M (2010). "**amei**: An R Package for the Adaptive Management of Epidemiological Interventions." *Journal of Statistical Software*, **36**(6), 1–32. URL http://www.jstatsoft.org/v36/i06/.

Murray JD (2002). *Mathematical Biology I: An Introduction*. Springer Verlag.

O'Neill P (2002). "A Tutorial Introduction to Bayesian Inference for Stochastic Epidemic Models Using Markov Chain Monte Carlo Method." *Mathematical Biosciences*, **180**, 103–114.

O'Neill P, Becker N (2001). "Inference for an Epidemic When Susceptibility Varies." *Biostatistics*, **2**(99–108).

O'Neill P, Roberts G (1999). "Bayesian Inference for Partially Observed Stochastic Epidemics." *Journal of the Royal Statistical Society, Series A*, **162**, 121–129.

Patel R, Longini I, Halloran M (2005). "Finding Optimal Vaccination Strategies for Pandemic Influenza Using Genetic Algorithms." *Journal of Theoretical Biolog*, **234**, 201–212.

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

Rotondi M (2007). ***epibasix***: *Elementary Epidemiological Functions for a Graduate Epidemiology/Biostatistics Course.* R package version 0.1.

Stevenson M, Nunes T, Sanchez J, Thornton R (2008). ***epiR***: *Functions for Analysing Epidemiological Data.* R package version 0.9-11, URL http://epicentre.massey.ac.nz.

Streftaris G, Gibson G (2004). "Bayesian Inference for Stochastic Epidemics in Closed Populations." *Statistical Modelling*, **4**, 63–65.

Tildesley M, Savill N, Shaw D, Deardon R, Brooks S, Woolhouse M, Grenfell B, Keeling M (2006). "Optimal Reactive Vaccination Strategies for a Foot-and-Mouth Outbreak in the UK." *Nature*, **440**, 83–86.

Zwane E (2007). ***stochasticGEM***: *R Package for Fitting Stochastic General Epidemic Models.* R package version 0.0-1.

Affiliation:

Daniel Merl
Department of Applied Mathematics and Statistics
University of California, Santa Cruz
Santa Cruz, CA 95064
USA
E-mail: merl1@llnl.gov

Leah R. Johnson
Ecology, Evolution, and Marine Biology
University of California, Santa Barbara
Santa Barbara, CA 93106
USA
E-mail: lrjohnson@lifesci.ucsb.edu
URL: http://www.lifesci.ucsb.edu/~lrjohnson/

Robert B. Gramacy
Statistical Laboratory
University of Cambridge
Wilberforce Road
Cambridge, CB3 0WB
United Kingdom
E-mail: bobby@statslab.cam.ac.uk
URL: http://www.statslab.cam.ac.uk/~bobby

Marc Mangel
Department of Applied Mathematics and Statistics
University of California, Santa Cruz
Santa Cruz, CA 95064
USA
E-mail: msmangel@ucsc.edu
URL: http://users.soe.ucsc.edu/~msmangel/