

1 **Title**

2 A Statistical Framework for the Adaptive Management of Epidemiological Interventions

3 **Authors and Affiliations**

4 Daniel Merl (corresponding author), Department of Statistical Science, Duke University.

5 Email: dan@stat.duke.edu

6 Leah R. Johnson, Department of Ecology, Evolution, and Marine Biology. University of

7 California, Santa Barbara.

8 Robert B. Gramacy, Statistical Laboratory, University of Cambridge

9 Marc Mangel, Center for Biomolecular Science and Engineering and Department of Applied

10 Mathematics and Statistics, University of California, Santa Cruz.

11 **Abstract/Summary**

12 **Background:** Epidemiological interventions aim to control the spread of infectious disease  
13 through various mechanisms, each carrying a different associated cost.

14 **Methodology:** We describe a flexible statistical framework for generating optimal  
15 epidemiological interventions that are designed to minimize the total expected cost of an  
16 emerging epidemic while simultaneously propagating uncertainty regarding the underlying  
17 disease model parameters through to the decision process. The strategies produced through this  
18 framework are adaptive: vaccination schedules are iteratively adjusted to reflect the anticipated  
19 trajectory of the epidemic given the current population state and updated parameter estimates.

20 **Conclusions:** Using simulation studies based on a classic influenza outbreak, we demonstrate the  
21 advantages of adaptive interventions over non-adaptive ones, in terms of cost and resource  
22 efficiency, and robustness to model misspecification.

## 23 **Introduction**

24 Epidemiological interventions generally remove susceptible individuals or apply some form of  
25 treatment to infected individuals in order to prevent further spread of a disease. The susceptible  
26 population may be culled, as in the case of foot-and-mouth disease [1,2], in which case the total  
27 population size is permanently reduced. The infected population may be quarantined, as in the  
28 case of SARS [3], in which case total population size is unchanged but the fraction of infecteds  
29 that may be in contact with susceptibles is reduced. Most commonly, susceptibles are vaccinated  
30 (cf influenza or smallpox [4,5]), in which case the total number of susceptibles, but not the total  
31 population size, is reduced.

32 Each of these interventions incurs a quantifiable cost: culling results in additional deaths;  
33 medical treatments or quarantines result in monetary expenses; vaccination incurs both monetary  
34 expenses, and in some cases additional vaccine-induced infections. Additionally, in many  
35 situations the costs associated with each of these actions can depend upon the state of the disease  
36 within the population of interest. For example, per-dosage prices of vaccine can increase as  
37 resources become scarce as a result of an aggressive vaccination campaign. Similarly, vaccine  
38 efficacy can decrease as a result of selection for drug resistance. Such observations raise the  
39 question of how to find optimal interventions that adaptively depend on the state of the epidemic.

40 A key challenge to calculating optimal intervention strategies involves devising ways to  
41 characterize and explore the space of intervention policies. Most existing work on optimal  
42 intervention has required various limiting assumptions about the forms of such strategies. Ball  
43 and Lyne [6] considered optimal vaccination in terms of the allocation of vaccine doses to  
44 households of various sizes in an explicitly structured population model. Patel *et al* [7]  
45 considered optimal vaccination in terms of the allocation of vaccine doses to different age classes

46 in an explicitly age- and geographically- structured population model. Tildesley *et al* [1]  
47 describe optimal vaccination strategies for a foot-and-mouth epidemic in which the optimized  
48 parameter is the size of the radius surrounding a point of infection within which all livestock are  
49 to be vaccinated. These methods are primarily concerned with pre-emptive interventions that  
50 can be completed before the arrival of the pathogen. Under such scenarios, there is no need to  
51 consider adaptive or sequentially updated interventions because as soon as the intervention  
52 policy is triggered, the threat of epidemic is eradicated. In real scenarios, such widespread  
53 vaccination may not be achievable. Moreover, these methods traditionally involve calculations  
54 that assume no uncertainty in key model parameters such as transmission rate, recovery rate, and  
55 mortality rate. Recently Elderd *et al* [8], using Bayesian methods, demonstrated the importance  
56 of explicitly quantifying such underlying uncertainty when forecasting the expected efficacy of  
57 trace versus mass vaccination policies. Their findings demonstrate that accurate propagation of  
58 parameter uncertainty can sometimes reveal deep and troubling consequences of a proposed  
59 vaccination strategy, and they suggest that incorporation of such uncertainty could impact policy  
60 decisions.

61 Here we address the question of how to dynamically propagate uncertainty in order to  
62 respond to an emerging epidemic while simultaneously and continuously learning about its  
63 underlying transmission dynamics. Estimation of model parameters is facilitated by regarding  
64 the transmission dynamics as stochastic processes rather than deterministic solutions to a  
65 structural equation model. This allows us to explicitly account for uncertainty in both model  
66 parameters and disease transmission. We consider a very general class of vaccination strategies  
67 defined by a fraction of the current susceptible population to be targeted for vaccination, and a  
68 threshold number of susceptibles such that once the number of susceptibles falls below this

69 threshold, the vaccination campaign is called off. We demonstrate the calculation and  
70 application of optimal strategies of this form when coupled with iteratively updated parameter  
71 estimates using simulations based on a well-studied influenza outbreak [9]. Our emphasis is not  
72 on the realism of the underlying SIR model (though it has been shown that even simplistic  
73 transmission models can provide good fit to actual data [10]), but rather to describe an effective  
74 approach for combining estimation and policy calculation. Permitting greater flexibility in the  
75 form of the possible intervention renders calculation of optimal intervention strategies  
76 analytically intractable, thus requiring evaluation by Monte Carlo-based methods. Once in a  
77 Monte Carlo-based framework, it becomes straightforward to couple the evaluation of  
78 intervention strategies with Bayesian procedures for performing on-line estimation of parameters  
79 of the underlying epidemic model, thereby propagating parameter uncertainty through to policy  
80 decisions.

81         The policies produced by this framework are optimal in that they minimize the expected  
82 cost of the epidemic and adaptive in that the optimal policy changes as a function of the state of  
83 the epidemic and the degree of uncertainty in underlying model parameters. Using extensive  
84 simulation studies we compare the distribution of costs accrued under adaptive intervention to  
85 those arising from non-adaptive policies in a variety of scenarios. Our studies show that adaptive  
86 policies perform similarly to nonadaptive policies based on perfect parameter estimates, and  
87 significantly better than nonadaptive policies based on imperfect parameter estimates.  
88 Additionally, we show that adaptive online estimation affords the method some robustness to  
89 model misspecification. These results further demonstrate the importance of accounting for such  
90 underlying uncertainties in dynamic settings and indicate the utility of adaptive policies in  
91 settings where perfect estimates and a true model do not exist. All computational methods used

92 herein have been made freely available through the `amei` (Adaptive Management of  
93 Epidemiological Interventions) R package [11].

## 94 **Results**

95 A classic study of Murray's [9] describes the spread of influenza through the population of a  
96 British boarding school. During the course of the epidemic, which was traced to the arrival of a  
97 single infectious student, all 763 students were eventually infected. The epidemic conforms to  
98 many standard assumptions of SIR models: a population essentially closed to immigration and  
99 emigration, recovery with immunity, and nearly homogeneous mixing of susceptibles and  
100 infectives.

101 Viewing the transmission dynamics as a discrete time stochastic process rather than a  
102 deterministic system of coupled differential equations implies a distribution of possible outcomes  
103 for the epidemic. By conditioning on parameter values and initial conditions ( $S_0 = 762, I_0 = 1$ ),  
104 Monte Carlo simulation can be used to explore the distributions of numbers of susceptible,  
105 infected, and recovered individuals, as well as total accrued cost, as functions of time. Murray  
106 provides estimates of the transmission rate ( $\beta = 0.00218$ ) and recovery rate ( $\gamma = 0.4$ ), which we  
107 regard as the "true" underlying parameter values in our simulations. Additional aspects of the  
108 transmission function are discussed in the Methods section. We assume that all costs can be  
109 expressed in a common monetary cost unit. Other choices of cost functions that address the  
110 issue of nonconformable costs (e.g. lives vs dollars) are mentioned in the Discussion.

111 Setting the unit cost to be that of maintaining a single infected individual for one time  
112 step (cost per infected,  $c_i = 1$ ), repeated forward simulation of the epidemic (Figure 1) indicates  
113 that the mean total cost over 40 time steps is approximately 2100 cost units (Figure 2),

114 attributable entirely to the cumulative cost of maintaining a large population of infected  
115 individuals until recovery.

116

117 **Variable Stop Time Vaccination.** We consider a relatively simple but flexible class of  
118 intervention strategies that involve vaccinating a target fraction ( $\alpha$ ) of susceptible individuals at  
119 each time step. After a round of vaccination, if the number of remaining susceptibles is less than  
120 a designated threshold ( $\gamma$ ), the vaccination campaign is discontinued. Policies defined in this  
121 way provide effective target population sizes, to which post-hoc corrections can be applied in  
122 light of knowledge of the population structure.

123 We assume that in a single time unit there is an upper bound on the maximum targetable  
124 fraction of susceptibles. In our simulations we set this bound to be 30%, so that several time  
125 units are required to vaccinate the majority of susceptibles. We also assume there is a period of  
126 time after the arrival of the initial infection before intervention can begin. In our examples, we  
127 assume this lag time to be 7 time units. These values are chosen purely for the purpose of  
128 demonstration, and can be assigned any value in the `amei` software.

129 The optimal variable stop time vaccination strategy can be found by searching the policy  
130 space (i.e. pairs of fractions-to-vaccinate  $\alpha$  and stopping thresholds  $\gamma$ ) for the policy that most  
131 frequently produces the lowest expected cost. The calculation of the optimal policy therefore  
132 explicitly accounts for uncertainty associated with the disease transmission and recovery  
133 processes (see Methods) under a given valuation of the model parameters. Assuming a value of  
134 2 cost units per dose of vaccine ( $c_v = 2$ ), we use Monte Carlo simulation to estimate the  
135 expected cost surface associated with variable stop time policies based on the true parameter  
136 values (Figure 3). The minimum expected cost is achieved under a policy of maximum (30%)

137 vaccination and a stopping threshold of 150 individuals. Repeated simulation of the epidemic  
138 under this policy shows that in the average case (dashed line), the policy amounts to 4 time units  
139 of maximum vaccination as soon as the initial lag is over (Figure 4). In situations where the  
140 number of susceptibles remaining after the lag is already below 150 individuals, no policy is  
141 implemented. The 95% central interval for the final distribution of total vaccine units dispensed  
142 is (339,581), representing variation in the total size of the epidemic at the time of the vaccination  
143 sweep, and the numbers of new infections after vaccination begins. Figure 5 shows the  
144 distribution of total costs accrued under this policy. After the end of the vaccination campaign,  
145 the uncertainty bands widen, representing variations in the costs associated with maintaining the  
146 remaining population of infected individuals until their natural recoveries. The mean total cost at  
147 time 40 is 1652 cost units, approximately a 21% reduction in total cost compared to no-  
148 intervention.

149

150 **Adaptive Management.** The intervention calculated in the previous section represents a gold-  
151 standard for this particular scenario because the vaccination strategy was calculated using the  
152 same parameter values and the same SIR model formulation as the simulated disease process. In  
153 most settings it will be natural to regard the transmission model parameters as unknowns to be  
154 estimated from incoming count data describing the sizes of the susceptible, infected, and  
155 recovered subpopulations. In this section we describe the procedure for performing adaptive  
156 management of an emerging epidemic, in which we account for parameter uncertainty and its  
157 impact on vaccination strategies.

158 An epidemic can be effectively summarized by the disease state of the population (*i.e.* the  
159 current numbers of susceptible and infected individuals) and by the SIR model parameters that

160 define the dynamics of transmission, death, and recovery. In adaptive management, the former is  
161 used to perform inference on the latter. Each time new data are collected, Markov chain Monte  
162 Carlo (MCMC) is used to sample from the current posterior distribution on model parameters.  
163 The optimal variable stop time strategy associated with each set of sampled parameter values is  
164 calculated, and the policy that most frequently minimizes the total expected cost (over all  
165 sampled parameter values) is enacted at the next time step. The fundamental difference between  
166 the adaptive policies calculated here and those calculated in the previous section is that here, the  
167 vaccination policy is a dynamic function of the current disease state and the current distribution  
168 of each parameter, whereas before, the policy was a static function of the initial disease state and  
169 the initial point estimate of each parameter.

170         The effectiveness of this approach can be similarly explored by repeated simulation of  
171 epidemics under adaptive management. As before, we assume an initial lag time of 7 time units  
172 before vaccination begins. Here we also introduce a cost associated with deaths ( $c_d = 4$ ). Even  
173 though the “true” model does not include mortality, the fitted model includes a mortality  
174 parameter ( $\mu$ ). This allows examination of the degree to which adaptive management strategies  
175 are robust to model misspecification.

176         Initial uncertainty regarding parameter values is expressed in the form of  
177 vague/noninformative prior distributions, as specified in the Methods. The choice of prior  
178 distributions in Bayesian models is of fundamental importance, and other possible choices are  
179 mentioned in the Discussion section. At each time step, the state of the epidemic is advanced  
180 one time step using the same “true” parameter values used in the previous section. Intervention  
181 strategies, however, are calculated based on the current parameter estimates.

182            Figures 6 and 7 show the distributions of susceptible, infected, recovered, and vaccinated  
183 individuals, and total accumulated costs for repeated simulation of the epidemic under adaptive  
184 management. These dynamics can be compared to those in Figures 4 and 5 in order to explore  
185 the effect of propagation of parameter uncertainty on efficacy of control measures. Compared to  
186 Figure 4, the central 95% region associated with the total number of vaccine units dispensed over  
187 the course of the intervention is more compact: (351, 536) with mean of 428 units for the  
188 adaptive policy versus (339, 580) with a mean of 442 units for the nonadaptive policy. The  
189 tighter bound about a smaller mean is due to the ability of the adaptive strategies to methodically  
190 diminish the vaccination campaign as a function of the epidemic state. This can be seen in  
191 Figures 8 and 9, which display the distributions of implemented vaccination strategies for each  
192 time step during the course of adaptive management. In the average case (dashed line), the  
193 maximum policy is enacted for 3 time steps, followed by a round of 20% vaccination. The  
194 uncertainty surrounding the implemented strategies indicates the degree to which the the  
195 adaptive policies are adjusted in light of data. In epidemics associated with the upper 97.5  
196 percentile of vaccination strategies (top solid line in Figures 8 and 9), the adaptive policy calls  
197 for 4 rounds of maximum vaccination followed by a round of 20% vaccination, followed by a  
198 final round of 5% vaccination. In this way, the adaptive nature of the interventions enables more  
199 efficient use of vaccine resources than achieved under nonadaptive policies.

200            The distribution of total cost associated with the adaptive intervention simulations  
201 (Figure 7) is essentially equivalent to the distribution of costs achieved under static intervention  
202 with perfect information (Figure 5), indicating that even the short period of data collection prior  
203 to action produces parameter estimates that are sufficient for accurate prediction of the disease  
204 dynamics. Figure 10 shows the final posterior distributions on the four model parameters

205 estimated from the data during one simulation of the epidemic under adaptive management. True  
206 values are indicated with a circle, mean values are indicated with an 'x', and the central 95%  
207 region of each distribution is shaded. The prior densities of each parameter for the same interval  
208 are shown in red. As mentioned above, the inference model is misspecified relative to the model  
209 being used to simulate the epidemic, in that the inference model includes a mortality parameter  
210 ( $\mu$ , see **Methods**), even though no deaths were observed in the simulated outbreaks. By  
211 coupling the policy calculations with an inference framework, the effect of such model  
212 misspecification appears to be reduced.

213         We can further demonstrate the utility of the adaptive approach in situations of more  
214 severe model misspecification. To do so, we construct a simulation experiment in which the  
215 inference model upon which the adaptive management is based is as described here, but in which  
216 the underlying transmission model through which new infecteds are generated is an entirely  
217 different, non-nested transmission model with a latent infective reservoir (see the `amei` vignette  
218 on CRAN [11] for details). This situation more closely resembles one that may be encountered  
219 in practice, where new infections are arising from an actual disease transmission process whose  
220 dynamics are at best approximated by any mathematical characterization. Table 1 compares  
221 summaries of the posterior distribution of cumulative cost arising under adaptive management to  
222 those predicted under the optimal static policy using parameters estimated for the misspecified  
223 model based on a completely observed epidemic. It is important to recognize that the adaptive  
224 policy is at a severe disadvantage, basing its actions on parameter estimates produced  
225 simultaneously during the course of a single epidemic (and using vague prior distributions) while  
226 the static policy conditions on parameter estimates obtained from a completely observed  
227 epidemic. In spite of this, the adaptive policy achieves nearly identical costs.

228           We have now shown the near equivalence of the adaptive and static policies in two  
229 different scenarios. These situations indicate that the proposed methodology is efficiently and  
230 with sufficient accuracy estimating the parameters of the transmission model, such that adaptive  
231 strategies based on these on-line estimates produce equivalent outcomes to those static strategies  
232 based on full retrospective analyses. Moreover, it is simple to demonstrate that static control  
233 measures based on reasonable but imperfect parameter estimates can lead to substantially worse  
234 outcomes/higher costs than the adaptive policies (Table 2). In real situations, where actions must  
235 be based on parameter estimates made from incomplete or limited information, the practice of  
236 iterative refinement of estimates and policies is likely to result in significantly improved  
237 outcome.

## 238 **Discussion**

239 We have demonstrated a novel adaptive management strategy based on a relatively simple  
240 characterization of the underlying SIR model and the epidemiological cost function. In  
241 principle, this methodological framework can readily accommodate more complicated disease  
242 dynamics such as immigration, latent infected states, missing data, and vector-communicated  
243 diseases, as well as more complicated intervention strategies that allow combined vaccination  
244 and quarantine. However, the incorporation of such features is likely to impose a heavy  
245 computational burden, and so model complexity should only be increased when additional  
246 parameters are supported (and identified) by the data and demanded by the biology. As in all  
247 Bayesian analyses, care must be taken when choosing prior distributions. In this study, our  
248 primary interests required the use of vague/noninformative prior distributions, in order to  
249 demonstrate the estimability of model parameters. In practice, informative, even pessimistic  
250 priors (i.e., overestimated infectiousness and mortality, underestimated recovery) may provide

251 useful reference points for the adaptive policy calculations, especially in situations of acute  
252 infections for which the duration of the epidemic may be too short for incoming data to dominate  
253 the prior information. In such situations, the adaptive approach still provides the opportunity for  
254 data to inform parameter values if it becomes available, while basing interventions on current  
255 parameter estimates as determined by their prior distributions.

256         There is an important choice to be made in assigning costs to the various actions that  
257 comprise an intervention strategy. A monetary valuation scheme is the most straightforward, but  
258 it may be difficult to construct such a scheme that adequately represents all aspects of the  
259 decision. One alternative would be a valuation in which each cost is chosen to represent a  
260 probability of mortality. In this way, the cost to be minimized would be the expected total loss of  
261 life for the epidemic under a given intervention strategy. By assuming that the removal rate can  
262 be expressed as  $\rho = \mu + \nu$ , where  $\mu$  is the rate of disease-induced mortality and  $\nu$  is the rate of  
263 natural recovery from the infected state, we can set  $c_i = (1 - e^{-\rho}) \frac{\mu}{\rho}$ , so that the cost associated  
264 with maintaining a given number of infected individuals for a unit of time is the number of  
265 infected individuals that are expected to die in a unit of time. Similarly, situations exist where it  
266 is reasonable to assign a probability of mortality to the removal of susceptibles, as in the cases of  
267 smallpox vaccination or the culling of livestock.

268         A related extension to this framework would involve applying a monetary constraint to a  
269 loss-of-life cost function. If we were to assume  $p_i$  and  $p_r$  to be, respectively, the probabilities of  
270 mortality associated with untreated infected individuals and the removal of susceptibles, and  
271 define  $d$  to be the monetary resources available for the intervention, then within this framework it  
272 is possible to find the intervention that minimizes the total loss-of-life subject to the total

273 spending constraint  $d$ . Similarly, it would be possible to optimize with respect to some selective  
274 criterion in order to preserve vaccine efficacy rather than select unnecessarily for drug-resistant  
275 pathogens. Also note the possibility of calculating policies based on minimization of some  
276 quantile of the realized cost rather than the mean cost. This would lead to minimization of costs  
277 associated with worst case scenarios, rather than that associated with the average case scenario.  
278 These and other alternative formulations of the underlying optimization problem can be easily  
279 accommodated in the framework presented here.

280         The utility of adaptive interventions is especially evident in situations of an emerging  
281 pathogen with which the host population has no previous experience. In such a situation,  
282 vaccines will not be immediately available at the onset of the epidemic, and so a methodology  
283 for combining currently available actions while anticipating the future availability of vaccines  
284 would be of great use. Effective epidemiological intervention requires swift decision in  
285 consideration of the various direct and indirect costs of intervention. The methodological  
286 framework described here provides a decision theoretic basis for automating this process.

## 287 **Materials and Methods**

288 All statistical and computational methodology described here has been implemented in a freely  
289 available R package called `amei` (Adaptive Management of Epidemiological Interventions),  
290 which can be downloaded at <http://cran.r-project.org/web/packages/amei/index.html> [11].

291  
292 **SIR Model.** We consider a standard Susceptible-Infected-Removed (SIR) model [10,12] with  
293 no loss of immunity but with mortality. In this model, the dynamic variables at time  $t$  are the  
294 number of susceptible individuals,  $S(t)$ ; the number of infected individuals,  $I(t)$ ; the number of  
295 recovered individuals,  $R(t)$ ; and the number of removed/dead individuals,  $D(t)$ . We assume that

296 the population is closed to immigration such that  $S(t)+I(t)+R(t)+D(t)=N$  is constant, and any  
297 three of the dynamic variables define the fourth.

298 To characterize the transmission of the disease, we adopt the negative binomial form for  
299 the transmission function [13], so that the model parameters are the transmission rate  $b$ , the  
300 overdispersion parameter  $k$ , the death rate  $\mu$ , and the rate of recovery to the immune class  $\nu$ .  
301 Under these assumptions, the SIR model is described by the following system of differential  
302 equations [12, 13]:

$$303 \quad \frac{dS}{dt} = -kS \log\left(1 + \frac{bI}{k}\right) \quad (1)$$

$$304 \quad \frac{dI}{dt} = kS \log\left(1 + \frac{bI}{k}\right) - (\nu + \mu)I \quad (2)$$

$$305 \quad \frac{dR}{dt} = \nu I \quad (3)$$

$$306 \quad \frac{dD}{dt} = \mu I \quad (4)$$

307 The negative binomial transmission function implies that disease transmission occurs following a  
308 Poisson process in which encounters between infected and susceptible individuals are Poisson  
309 distributed with the encounter rate varying according to a gamma distribution with coefficient of  
310 variation  $k^{-1/2}$ . Via the parameter  $k$ , the negative binomial transmission function can account for  
311 social interactions and/or network factors in disease transmission, without requiring explicit  
312 characterization of the population structure.

313 The SIR model formulation also leads immediately to a natural discrete time  
314 approximation for the numbers of infections ( $\tilde{I}$ ), recoveries ( $\tilde{R}$ ) and deaths ( $\tilde{D}$ ) arising in the

315 unit time interval from  $t$  to  $t+1$ . Holding the total number of infected individuals  $I$  constant and  
 316 integrating Equation 1 over a unit time interval gives

$$317 \quad S(t+1) = S(t) \left[ \frac{k}{k + bI(t)} \right]^k \quad (5)$$

318 so that the fraction of susceptible individuals surviving a unit time interval is  $\left[ \frac{k}{k + bI(t)} \right]^k$ .

319 Viewed as a discrete time stochastic process, the number of new infections occurring between  
 320 time  $t$  and  $t+1$  when  $S(t)=s$  and  $I(t)=i$  can be described by

$$321 \quad \tilde{I} | s, i \sim \text{Bin}(s, p_i(i, b, k)) \quad (6)$$

322 where  $p_i(i, b, k) = 1 - \left( \frac{k}{k + bi} \right)^k$  and  $\text{Bin}(n, \pi)$  is the standard binomial distribution. Similarly, by

323 integrating Equations 3 and 4, we have that the numbers of recoveries and deaths occurring  
 324 between time  $t$  and  $t+1$  can be described by

$$325 \quad \tilde{R} \sim \text{Bin}(i, p_r) \quad (7)$$

$$326 \quad \tilde{D} \sim \text{Bin}(i - \tilde{r}, p_d) \quad (8)$$

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

$$329 \quad S(t+1) = S(t) - \tilde{I} | s, i \quad (9)$$

$$330 \quad I(t+1) = I(t) - \tilde{R} | i - \tilde{D} | i, \tilde{r} + \tilde{i} \quad (10)$$

331 Here lower case denotes the realized value of the associated capital letter random variable. In this  
 332 discrete time approximation we have assumed a particular ordering of events, namely that

333 recoveries occur first, followed by deaths from among those infected individuals who did not  
334 recover, followed by new infections. Simulation studies indicated that these assumptions, as well  
335 as other possible orderings, resulted in system dynamics that were approximately equal in  
336 expectation to deterministic solutions of the continuous time SIR model.

337         In all forward simulations of the disease dynamic (except where noted) we assume the  
338 “true” underlying parameter values to be those estimated by Murray [9], with the exception of  
339 the negative binomial overdispersion parameter  $k$ . Thus,  $b=0.00218$ ,  $v=0.4$ , and  $\mu=0$  (no disease-  
340 related mortality). We set the overdispersion parameter to be  $k=0.1$ , in order to produce  
341 epidemics that, without intervention, have run their course by 40 time units but such that there is  
342 variation in the size of the outbreak.

343

344 **Epidemiological Cost Function.** We formulate the total expected cost of the epidemic in terms  
345 of the underlying costs associated with maintaining infected individuals until recovery, suffering  
346 death, and administering vaccinations. Let  $c_1(\alpha, \gamma, s)$  denote the cost associated with  
347 interventions involving susceptibles when  $S(t)=s$ . Here  $\alpha$  is the fraction of susceptibles that are  
348 moved directly into an immune/recovered class, as by vaccination, and  $\gamma$  is the threshold below  
349 which the intervention is discontinued. Letting  $c_v$  denote the cost per unit, then

$$350 \quad c_1(\alpha, \gamma, s) = \begin{cases} c_v \alpha s & \text{if } s > \gamma \\ 0 & \text{if } s \leq \gamma \end{cases} \quad (11)$$

351 We let  $c_2(i)$  denote the cost associated with interventions involving infecteds when  $I(t)=i$ . This  
352 component includes the costs associated with maintaining the non-recovered infected individuals  
353 and costs associated deaths, as in

354 
$$c_2(i) = c_i i + c_d \tilde{d} \tag{12}$$

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

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

359 
$$E\{C_0\} = c_1(\alpha, \gamma, s_0) + c_2(i_0) + E\{C_1\} \tag{13}$$

360 where  $E\{C_t\}$  denotes the expected cost accumulated from time  $t$  onwards. The optimal  
361 intervention strategy  $(\alpha, \gamma)$  is the one that minimizes the total accumulated cost over the course of  
362 the epidemic. Two methods for calculating such strategies are as follows.

363  
364 **Calculating Variable Stop Time Vaccination Strategies.** The total expected cost depends on  
365 the parameter values and the initial epidemiological state  $(s_0, i_0)$ . Thus, conditional on a set of  
366 parameter values, Monte Carlo simulation can be used to search over values of  $\alpha$  and  $\gamma$  in order  
367 to find the combination that minimizes  $E\{C_0\}$ . For each combination of  $\alpha$  and  $\gamma$ , with  $\alpha$  ranging  
368 from 0 to 0.7 and  $\gamma$  from 0 to 750 in increments of 50, we conduct 100 simulations of the  
369 epidemic, using the true parameter values, in order to estimate the mean cost associated with the  
370 intervention  $(\alpha, \gamma)$ . The strategy producing the lowest mean cost is defined to be the optimal  
371 intervention.

372  
373 **Calculating Adaptive Management Strategies.** As above, the expected cost surface associated  
374 with a given set of parameter values (as obtained by MCMC, described below), can be explored

375 using standard Monte Carlo methods. At each time step, MCMC is used to produce 10,000  
376 samples from the current posterior distribution on model parameters. These samples are thinned  
377 to 100 samples, and for each of these 100 samples the optimal variable stop time vaccination  
378 strategy is calculated as described above. The adaptive strategy to be implemented at that time  
379 step is defined to be the most frequently optimal strategy for the 100 posterior samples.

380 Notice that if we were to allow the fraction of the population targeted for vaccination to  
381 be a function of future disease states, rather than a static fraction and a stopping threshold, we  
382 could regard Equation 13 as a stochastic iteration equation and use stochastic dynamic  
383 programming [14] to calculate the optimal intervention associated with a set of parameter values.  
384 Such an approach may be useful for situations in which knowledge of the disease state is  
385 available, but for whatever reason sequential inference is not possible. In the situation considered  
386 here, in which the static strategy is sequentially updated based on the current disease state and  
387 parameter estimates, the adaptive strategy that emerges is similarly flexible, in that it consists of  
388 a state-dependent sequence of target fractions, but does not involve the additional computational  
389 burden associated with stochastic dynamic programming.

390

391 **Online Parameter Estimation.** We use Markov Chain Monte Carlo (MCMC) [15] to learn  
392 about the posterior distribution of  $b$ ,  $k$ ,  $v$ , and  $\mu$  conditioned on the evolution of the epidemic  
393 observed so far. The likelihood is given recursively in Equations 6—8. Assume, at first, that no  
394 intervention strategy is implemented. Let  $\tilde{i}_t = S(t-1) - S(t)$  be the number of new infecteds at  
395 time  $T$ , and similarly for the newly recovered and dead individuals  $\tilde{r}_t$  and  $\tilde{d}_t$  so that  
396  $\tilde{r}_t + \tilde{d}_t \leq I(t-1)$ . Then, the likelihood up to  $T$  is given by

397 
$$p(\{\tilde{i}\},\{\tilde{r}\},\{\tilde{d}\}|\dots) = \prod_{t=1}^T \text{Bin}(\tilde{i}_t | S(t-1), p(I(t-1), b, k))$$

398 
$$\prod_{t=1}^T \text{Bin}(\tilde{r}_t | I(t-1), p_r) \prod_{t=1}^T \text{Bin}(\tilde{d}_t | I(t-1) - \tilde{r}_t, p_d)$$

399 (14)

400 and we can see that it consists of three mutually independent components.

401 Conditional conjugacy can be exploited for  $\nu$  and  $\mu$  via Beta priors for  $p_r$  and  $p_d$ . A

402 Beta( $\alpha_r, \beta_r$ ) prior for  $p_r$  implies that

403 
$$p(\nu) = (1 - e^{-\nu})^{\alpha_r - 1} e^{-\nu \beta_r}$$

404 (15)

405 Conjugate updating leads to the posterior conditional

406 
$$p_r | \dots \sim \text{Beta}(\alpha_r + \sum_{t=1}^T \tilde{r}_t, \beta_r + \sum_{t=1}^T (I(t) - \tilde{r}_t))$$

407 (16)

408 The form of the conditional posterior for  $\nu$  is similar to Equation 16 and can be simulated by first

409 drawing  $p_r$  via Equation 16 and then applying the inverse transformation  $\nu = -\log(1 - p_d)$ .

410 Sampling for  $\mu$  proceeds similarly with

411 
$$p_d | \dots \sim \text{Beta}(\alpha_d + \sum_{t=1}^T \tilde{d}_t, \beta_d + \sum_{t=1}^T (I(t) - \tilde{r}_t - \tilde{d}_t))$$

412 (17)

413 So it is possible to take Gibbs samples for  $\nu$  and  $\mu$  so long as appropriate  $\alpha_r, \beta_r, \alpha_d, \beta_d$  can be

414 found to represent our prior beliefs. In ignorance we simply set these to unity, leading to uniform

415 priors on  $p_r$  and  $p_d$ .

416 Obtaining samples for  $b$  and  $k$  requires the Metropolis–Hastings algorithm. Our prior

417 beliefs can be encoded with gamma distributions, and conditional on a previous sample  $(b, k)$  the

418 next sample  $(b', k')$  can be obtained by Metropolis-within-Gibbs steps using:

414 
$$p(b' | k, \dots) \propto \Gamma(b' | \alpha_b, \beta_b) \prod_{t=1}^T \text{Bin}(\tilde{i}_t | S(t-1), p_t(I(t-1), b', k)) \quad (18)$$

415 
$$p(k' | b', \dots) \propto \Gamma(k' | \alpha_k, \beta_k) \prod_{t=1}^T \text{Bin}(\tilde{i}_t | S(t-1), p_t(I(t-1), b', k')) \quad (19)$$

416 For the prior settings, we currently use  $(\alpha_b, \beta_b) = (\alpha_k, \beta_k) = (1, 3)$  which (though seemingly  
 417 informative at first glance) turns out to be uninformative on the scale of the support of the posterior.  
 418 We find that random walk uniform proposals on the positive real line, i.e.,  $b' \sim U[3b/4, 4b/3]$ , gives  
 419 reasonably good mixing from the Markov chain, as evidenced by visual inspection of parameter  
 420 traces and other convergence diagnostics. More details pertaining to technical issues such as  
 421 MCMC convergence appear in the `amei` vignette [11].

422 The presence of a vaccination strategy necessitates a simple change to the above  
 423 equations. Replace  $S(t-1)$  with  $S(t-1) - v_t$ , where  $0 \leq v_t \leq S(t-1)$  is the number of susceptibles  
 424 which have been vaccinated. Then  $\tilde{i}_t = S(t-1) - v_t - S(t)$ .

## 425 Acknowledgments

426 We thank an associate editor and two anonymous referees for comments that greatly improved  
 427 the manuscript.

## 428 References

- 429 1. Tildesley, M. J., Savill, N. J., and Shaw, D. J. et al (2006) Nature 440, 83-86.  
 430 2. Enserink, M. (2001) Rapid response could have curbed foot-and-mouth epidemic.  
 431 Science 294: 26-27.

- 432 3. Lloyd-Smith, J. O., Galvani, A. P., and Getz, W. M. (2003) Curtailling transmission of  
433 severe acute respiratory syndrome within a community and its hospital. Proceedings of  
434 the Royal Society of London 270: 1979-1989.
- 435 4. Ferguson, N. M., Keeling, M. J., Edmunds, W. J., Gani, R., Grenfell, B. T., et al (2003)  
436 Planning for smallpox outbreaks. Nature 425: 681-685.
- 437 5. Halloran, M. E., Longini, I. M., Nizam, A., and Yang, Y. (2002) Containing bioterrorist  
438 smallpox. Science 298: 1428-1432.
- 439 6. Ball, F., and Lyne, O. (2002) Optimal vaccination policies for stochastic epidemics  
440 among a population of households. Mathematical Biosciences 177 & 178: 333-354.
- 441 7. Patel, R., Longini, I. M., and Halloran, M. E. (2005) Finding optimal vaccination  
442 strategies for pandemic inuenza using genetic algorithms. Journal of Theoretical Biology  
443 234: 201-212.
- 444 8. Elderd, B., Dukic V., and Dwyer G. (2006) Uncertainty in prediction of disease spread  
445 and public health responses to bioterrorism and emerging diseases. Proc Nat Acad Sci  
446 USA 103: 15693-15697.
- 447 9. Murray, J. D. (2002) Mathematical Biology I: An Introduction. Springer Verlag.
- 448 10. Anderson, R. M., and May, R. M. (1991) Infectious Diseases of Humans: Dynamics and  
449 Control. Oxford University Press.
- 450 11. Daniel Merl, Leah R. Johnson, Robert B. Gramacy and and Marc S. Mangel (2008).  
451 *amei*: Adaptive Management of Epidemiological Interventions. R package version 1.0.  
452 <http://cran.r-project.org/web/packages/amei/index.html>
- 453 12. Hetchote, H. W. (2000) The mathematics of infectious diseases. SIAM Review 42:599-  
454 653.

- 455 13. McCallum, H., Barlow, N., and Hone, J. (2001) How should pathogen transmission be  
456 modelled? Trends in Ecology and Evolution 16: 295-300.
- 457 14. Clark, C. W. and Mangel, M. (2000) Dynamic State Variable Models in Ecology:  
458 Methods and Applications. Oxford University Press.
- 459 15. Gamerman, D. and Lopes, H. (2006) Markov Chain Monte Carlo: Stochastic Simulation  
460 for Bayesian Inference. Chapman & Hall/CRC.

## 461 **Figure Legends**

462 Figure 1: Simulated epidemics. (2.5,50,97.5)-% quantiles for numbers of susceptible, infected,  
463 and recovered individuals over 1000 simulations of the epidemic without intervention.

464

465 Figure 2: Expected costs under nonintervention. (2.5, 50, 97.5)-% quantiles for total cost  
466 accrued over 1000 simulations of the epidemic without intervention. The mean total cost after 40  
467 days is 2100 cost units, with quantile bounds (1949,2263).

468

469 Figure 3: Expected cost surface for static interventions. The heatmap depicts the expected cost  
470 surface associated with variable stop time vaccination strategies based on the true parameter  
471 values. The minimum expected cost (1640 cost units) is achieved by a strategy of vaccinating  
472 30% of susceptibles at each time step, until the number of susceptibles falls below 150. The  
473 maximum expected is realized through inaction (top row and left column policies are never  
474 implemented).

475

476 Figure 4: Simulated epidemics under static intervention. (2.5, 50, 97.5)-% quantiles for the  
477 numbers of susceptible, infected, recovered, and vaccinated individuals over 1000 simulations of

478 the epidemic under the optimal variable stop time strategy based on true parameter values. The  
479 mean number of vaccine units dispensed is 442, with quantile bounds (339,580).

480

481 Figure 5: Expected costs under static intervention. Costs under optimal (2.5, 50, 97.5)-%  
482 quantiles for the total cost accrued over 1000 simulations of the epidemic under the optimal  
483 variable stop time strategy based on true parameter values. The mean total cost is 1652 cost  
484 units, with quantile bounds (1440,1846).

485

486 Figure 6: Simulated epidemics under adaptive management. (2.5, 59, 97.5)-% quantiles for the  
487 numbers of susceptible, infected, recovered, and vaccinated individuals over 100 simulations of  
488 the epidemic under optimal adaptive management. The mean number of vaccine units dispensed  
489 is 428, with quantile bounds (351,536).

490

491 Figure 7: Expected costs under adaptive management. (2.5,50,97.5)-% quantiles for total cost  
492 accrued over 100 simulations of the epidemic under optimal adaptive management. The mean  
493 total cost is 1665 cost units, with quantile bounds (1450,1888).

494

495 Figure 8: Vaccination levels under adaptive management. (2.5, 50, 97.5)-% quantiles for the  
496 fraction of susceptibles vaccinated at each time step over 100 simulations of the epidemic under  
497 optimal adaptive management.

498

499 Figure 9: Stopping times under adaptive management. (2.5,50,97.5)-% quantiles for the policy  
 500 stop time at each time step over 100 simulations of the epidemic under optimal adaptive  
 501 management.

502

503 Figure 10: Online parameter estimates. Final posterior density estimates for the transmission rate  
 504 (A), overdispersion parameter (B), recovery rate (C), and mortality rate (D). “True” parameter  
 505 values are indicated by a dot, mean posterior values are indicated by an ‘x’, and the central 95%  
 506 region of the distribution is shaded. Prior densities on the same regions are shown in red.

507

508 **Tables**

509

	2.5%-ile	Mean	Median	97.5%-ile
Adaptive	1910	2091	2089	2311
Nonadaptive	1888	2085	2085	2295

510 Table 1. Expected costs under model misspecification. Comparison of adaptive and  
 511 nonadaptive policy costs when the inference model is misspecified. Even though the static  
 512 policy is based on parameter estimates obtained after a completely observed epidemic, the costs  
 513 associated with adaptive management are similar.

514

	2.5%-ile	Mean	Median	97.5%-ile
Adaptive	1451	1665	1657	1888
Nonadaptive	1938	2103	2100	2264

515 Table 2. Expected costs under imperfect parameter estimates. Comparison of adaptive and  
516 nonadaptive policy costs when static management is based on imperfect parameter estimates  
517 ( $b = 0.001, \nu = 0.9, k = 10, \mu = 0$ ).



















