

Optimizing the control of disease infestations at the landscape scale

Graeme A. Forster[†] and Christopher A. Gilligan

Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom

Edited by Burton H. Singer, Princeton University, Princeton, NJ, and approved January 4, 2007 (received for review September 19, 2006)

Using a contact-process model for the spread of crop disease over a regional scale, we examine the importance of the time scale for control with respect to the cost of the epidemic. The costs include the direct cost of treating infected sites as well as the indirect costs incurred through lost yield. We first use a mean-field approximation to derive analytical results for the optimal treatment regimes that minimize the total cost of the epidemic. We distinguish short- and long-term epidemics, and show that seasonal control (short time scale) requires extreme treatment, either treating all sites or none or switching between the two at some stage during the season. The optimal long-term strategy requires an intermediate level of control that results in near eradication of the disease. We also demonstrate the importance of incorporating economic constraints by deriving a critical relationship between the epidemiological and economic parameters that determine the qualitative nature of the optimal treatment strategy. The set of optimal strategies is summarized in a policy plot, which can be used to determine the nature of the optimal treatment regime given prior knowledge of the epidemiological and economic parameters. Finally, we test the robustness of the analytical results, derived from the mean-field approximation, on the spatially explicit contact process and demonstrate robustness to implementation errors and misestimation of crucial parameters.

contact process | economic modeling | epidemiological modeling

There is increasing interest in the optimization of disease control at the landscape scale (1–5). This interest is motivated for crop and animal disease by the desire to reduce unnecessary use of pesticide and to minimize the cost of control relative to the costs of infection. Seeking an optimal strategy in this way opens up a rich suite of possibilities that range from treating all infected individuals, through treatment of a fixed or varying proportion of infected individuals, as circumstances change during the course of an epidemic, to no treatment at all. Finding the optimal strategy depends on the balance of economic and epidemiological parameters that reflect the nature of the host–pathogen system and the efficiency of control method. Progress therefore requires a combination of epidemiological and economic factors in modeling that hitherto have tended to remain separate.

Economic models have given insight into optimal control under constraints imposed by limited resources (6–8), but they frequently ignore the spatial and temporal dynamics of disease (9). Epidemiological modeling has largely focused on identifying the mechanisms responsible for epidemics (5, 10, 11) but has taken little account of economic constraints in analyzing control strategies. Recent exceptions include epidemiological models for the spread of severe acute respiratory syndrome (12, 13), foot and mouth disease (1, 2, 14–16), and avian flu (3), in which the efficiencies of different control strategies are compared, usually with respect to times for extinction of the invading pathogen. These models are often closely tied to the details of the specific host–pathogen system. Although these details are essential for the implementation of control, they sometimes obscure underlying principles of optimal disease control. A more generic approach applicable to a wide range of host–pathogen systems

is given by Dybiec *et al.* (17), who showed how matching the scale of the control strategy with the natural scale of the epidemic spread can bring epidemics under control, even when there is long-distance, small-world dispersal, as well as local dispersal of the pathogen. Much epidemiological work is predicated on the assumption that an invading disease must be eliminated. For many plant, animal, and even human diseases, however, it is not axiomatic that elimination of the pathogen is economically viable, which refocuses attention on the relative costs and benefits of control and the effects of epidemiological parameters associated with disease transmission and efficiency of control on them (18–21). One natural way to approach the situation, which allows analytical progress for simple epidemiological models, is through variational calculus (22, 23). Control strategies are introduced into epidemiological models, and optimal strategies are derived by minimizing an objective function that incorporates costs of infection and control. Sethi and others (19, 24–29) have successfully examined a number of these models. Here, we use the approach of Sethi (21), which utilized Green's theorem to solve the optimization problem, and build on it to examine the trade-offs between epidemiological and economic parameters. We motivate the analysis for the control of disease spreading through an agricultural landscape, by using the contact process (30) to describe an epidemic propagating through a lattice of susceptible sites. This type of model can be used to describe the spread of disease in a fixed landscape, where each site represents a bounded subpopulation that the disease can inhabit, such as a field. The analyses are principally motivated for crop disease, in which a field, plantation, or sometimes a farm is the natural epidemiological unit for monitoring and controlling the spread of disease through the landscape. Thus for a given crop, such as wheat, a field is initially classified as susceptible (S). It becomes infected (more strictly infectious) (I), as it becomes capable of transmitting infection to neighboring fields. Treatment of an infected field by a fungicide speeds the rate of recovery of the field to the susceptible class, and hence the system is described by an SIS model (31). This form of transmission is typical of a wide range of diseases that spread predominantly over short distances so that spread occurs between neighboring fields. The model applies to many soil-borne and splash-dispersed fungal pathogens and to viral pathogens vectored by short-range movement of aphid and other invertebrate vectors. By keeping the model generic, in this way the results also hold for simplified models of animal diseases, where the epidemiological unit is the herd, flock or, again, an entire farm.

Author contributions: G.A.F. and C.A.G. designed research; G.A.F. performed research; and G.A.F. and C.A.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviation: SIS model, susceptible-infected-susceptible to reinfection model.

[†]To whom correspondence should be addressed. E-mail: gaf32@cam.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0607900104/DC1.

© 2007 by The National Academy of Sciences of the USA

We show how the optimal control strategies depend on the balance between economic and epidemiological parameters as well as the initial density of infected units (e.g., fields) in the population. Specifically, we ask the following questions. Is the conventional strategy associated with the contact process optimal, whereby a fixed proportion of infected fields is treated? If not, what is optimal and why? We give particular attention to the importance of the time frame over which control is envisaged and how it affects the optimal strategy. A distinction is made between a short period, typically a single season, and a long time frame, extending over many seasons that would be typical of a perennial crop or of disease in natural environments. Analysis of strategies over a long term necessarily involves a discount rate to account for alternative investment of resources for control, whereas that for a single season arguably does not (23). We first derive analytical solutions from a mean-field approximation of the contact process, from which to identify optimal control strategies. These strategies are nonspatial; that is, they do not rely on any knowledge of the location of disease in the landscape. The advantage of such strategies is that they are less susceptible to implementation errors arising from cryptic infection, in which infectious sites remain nonsymptomatic for a period (18), than are strategies that track the explicit location of infections in the landscape. From these nonspatial strategies we derive policy plots that serve to identify optimal strategies, given some knowledge of the epidemiological and economic parameters and the current level of infestation. We then run extensive numerical simulations to test that these strategies are also optimal (among the set of all nonspatial strategies) for the spatially explicit contact process and therefore provide a link between the mean-field analysis and the motivational example (the spatial spread of disease). Finally, we consider the robustness of the strategies to implementation errors and misestimation of crucial parameters, and the advantages and limitations of this approach for the control of disease in the agricultural landscape are discussed.

The Model. In a contact process, each site of a lattice exists in one of two states, healthy or infected. Infected sites pass disease on to healthy neighbors and recover at a given rate, whereupon they become immediately susceptible to reinfection (SIS). Because a susceptible site must have an infected neighbor to become infected, the disease-free state is absorbing; persistence of the epidemic depends on the transmission rate and the recovery rate. By treating the occupancy of each site as statistically independent and assuming spatial homogeneity, we can derive a mean-field approximation for the proportion of infected sites (ρ) given by

$$\frac{d\rho}{dt} = -\epsilon_1\rho + \beta\rho(1 - \rho), \quad [1]$$

where ϵ_1^{-1} is the infectious period of each site and β the transmission rate, which is essentially an aggregated parameter composed of the transmission rate between pairs of infected and susceptible sites and the coordination number that describes the contact structure between infected and susceptible sites (30, 32).

The application of control to a given site may be assumed to reduce the infectious period leading to increased recovery rates from ϵ_1 to ϵ_2 , of that site to the susceptible state. By applying control randomly to k of the n possible sites in the landscape we derive the state equation with control $k \in [0, n]$, given by,

$$\frac{d\rho}{dt} = -\frac{k}{n}\epsilon_2\rho - \frac{(n-k)}{n}\epsilon_1\rho + \beta\rho(1 - \rho). \quad [2]$$

This equation implies that there exists a critical level of treatment $k = k_c$ given by

$$k_c = n \frac{\beta - \epsilon_1}{\epsilon_2 - \epsilon_1}, \quad [3]$$

such that if $k > k_c$ the disease will be eradicated in the long term. The existence of the phase transition is ensured by imposing the conditions $\epsilon_1 < \beta$ and $\epsilon_2 \geq \beta$. Note that although it is convenient to express the phase transition in terms of k_c because it maps onto control strategies as envisaged here, the phase transition could also be expressed in terms of the familiar epidemiological concept of the basic reproductive number, R_0 (10), by a simple transformation of Eq. 2 whereby $R_0 = \beta n / [(n - k)\epsilon_1 + k\epsilon_2]$. Thus, the condition $k > k_c$ is equivalent to $R_0 < 1$, the widely used epidemiological criterion for the prevention of disease invasion. The existence of a phase transition has also been proven for the contact process (33, 34). However, the critical parameter values at which the phase transition takes place differ from the mean-field case because of the effect of spatial structure. Approximations of the critical values have been provided by using simulations (35) and series analysis (36–39) for lattices with low dimensionality.

Optimal Control. The objective is to control the number of sites receiving treatment [$k = k(t)$] such that the total cost of infection and treatment is minimized over time. Under the assumption that the cost of treatment is linearly related to the amount administered, and similarly for the cost of infection (21, 29), the cost functions for short and long time scales are given, respectively, by

$$J_T = \int_0^T \left(c\rho + d\frac{k}{n} \right) dt, \quad J_\infty = \int_0^\infty e^{-rt} \left(c\rho + d\frac{k}{n} \right) dt, \quad [4]$$

where c is the cost per infected site per unit of time, d is the cost per treated site per unit of time, r is the discount rate, and T is the length of the season. The discount rate represents the rate the policymaker is willing to pay to trade off the value of treatment today against the ensuing cost of increased infection in the future (23). By using the Pontryagin maximum principle [supporting information (SI) Appendix A], the optimal solution for both the long and short time scale problems is given by

$$k = \bar{k} = \begin{cases} 0, & \text{if } f(t) > 0 \\ \in [0, n], & \text{if } f(t) = 0 \\ n, & \text{if } f(t) < 0, \end{cases} \quad [5]$$

where $f(t) = \psi\rho(\epsilon_2 - \epsilon_1) + d$ and ψ is the costate variable (or shadow price of infection). The costate variable is governed by the differential equation

$$\dot{\psi} = r\psi + c - \psi \left(-\frac{k}{n}\epsilon_2 - \frac{(n-k)}{n}\epsilon_1 + \beta(1 - 2\rho) \right). \quad [6]$$

Treatments applied to all or no sites ($k = n$ or $k = 0$) are termed extreme solutions, and intermediate coverage is termed a singular (or interior) solution. The costate variable indicates the marginal benefit of reducing the stock of infection. Because infection is harmful, it is always negative. A change Δk in the level of treatment causes the density of infected sites to change by $-\frac{\Delta k}{n}\rho(\epsilon_2 - \epsilon_1)$, and therefore the social value of this change is $-\psi\frac{\Delta k}{n}\rho(\epsilon_2 - \epsilon_1)$. The total cost of achieving such an outcome is $d\frac{\Delta k}{n}$. When treatment coverage is intermediate, so that variation in either direction is possible, the marginal cost of changing the coverage must be exactly equal to the social benefit. Thus, we obtain $-\rho\psi(\epsilon_2 - \epsilon_1) = d$, which is the condition $f(t) = 0$ for an interior (or singular) solution in Eq. 5. If the marginal cost of changing the coverage is different from the benefit received from the resulting change in infection, extreme coverage should be adopted. For example, if the cost of increasing coverage is less than the associated benefits, complete coverage

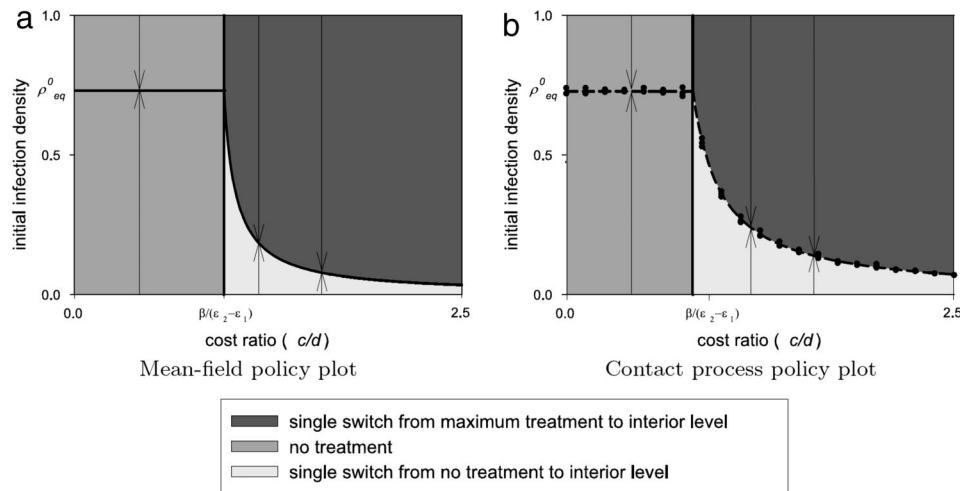


Fig. 1. Policy plots. (a) Mean-field model. (b) Contact process. Given that the initial density of infection is known, and we have an estimate of the epidemiological and economic parameters, we can use the policy plot as a tool to obtain the qualitative nature of the optimal treatment regime. The arrows indicate the progression toward the steady state for each strategy. In b, the dots represent the optimal equilibrium levels of infection for high, medium, and low initial infection densities. They vary because different epidemic realizations were used in each case. The three parameters are given in Table 1 and were used in both deterministic and spatially explicit cases.

($k = n$) should be adopted, which is equivalent to the condition $-\rho\psi(\epsilon_2 - \epsilon_1) > d$, or $f(t) < 0$ in Eq. 5 (40). The behavior of the function $f(t)$ dictates the nature of the optimal strategy. It is permissible to have solutions contained within the limits of treatment (singular solutions) only if $f(t) = 0$ for some nonzero length of time. Otherwise, the solution contains only the extreme levels of treatment, which is termed a “bang-bang” solution. A bang-bang solution may contain any number of switches from one extreme to the other over the course of the epidemic.

Results

The key difference between optimal short- and long-term disease control is the existence of a singular solution. It can be shown analytically (SI Appendix B) that singular solutions are not possible in the short time scale problem unless $\beta/(\epsilon_2 - \epsilon_1) = c/d$. This strict and exact condition is unlikely to be satisfied in natural systems, and so it can effectively be ruled out as a realistic option for policy decisions. Therefore, the optimal treatment regime involves extreme policies only, either treating all of the sites or none of them. Hence, the critical value k_c can be ruled out as a candidate. It can also be demonstrated analytically that if $\beta/(\epsilon_2 - \epsilon_1) < c/d$, then the optimal strategy is to treat all sites initially and then switch to no treatment at some determinable time (SI Appendix B). Otherwise, it is optimal not to treat at all.

For long-term control, the singular solution exists and is given by

$$k^* = k_c \left(1 + \frac{r\beta}{(\beta - \epsilon_1)(\beta - c/d(\epsilon_2 - \epsilon_1))} \right). \quad [7]$$

As in the case of short-term control, the qualitative nature of the optimal control strategy can be derived entirely analytically by analysis of the phase space (SI Appendix C and SI Fig. 4). There are three possibilities for the optimal path, which are summarized in a policy plot (Fig. 1a and SI Fig. 4). The optimal strategy is determined by a relationship that relates the economic and epidemiological parameters and also depends on the prevalence of the infection when treatment is first administered. There exists a critical relation such that when $c/d < \beta/(\epsilon_2 - \epsilon_1)$, the optimal strategy is to treat no sites at all, but when $c/d > \beta/(\epsilon_2 - \epsilon_1)$, the optimal strategy changes radically to a two-phase strategy that begins with treating all (or no) sites (depending on the initial density of infection) and then switches to the interior level of treatment. The optimal time to

switch is when the density of infection reaches the equilibrium value corresponding to $k = k^*$ (Fig. 1a). As the cost per infection increases relative to the cost per treated site, the optimal long-term level of infection approaches zero as the singular solution k^* asymptotes to k_c . Fig. 2 shows some illustrative examples of optimal long- and short-term treatment regimes and the corresponding paths of infection.

Numerical Testing and Sensitivity. With the qualitative nature of the optimal treatment strategies for the mean-field model established, we now revert to the original contact-process formulation to test these strategies. The optimal strategies developed by using the mean-field formulation are independent of space. They are effective, and arguably achievable strategies that are easily implemented in practice (providing we have an estimate of the disease parameters) because they do not rely on any detailed knowledge of the location of disease in the landscape. When we test these strategies on the contact process, we are not testing for absolute optimality over all strategies (because absolute optimality is sure to depend on a knowledge of spatial location of infection). We are testing for optimality over all nonspatial strategies; that is, strategies that do not rely on any knowledge of the location of disease (as in the mean-field case). Given that the mean-field approximation has the same qualitative properties as the explicitly spatial system, we use spatial simulation to test the hypothesis that the optimal treatment regimes also share the same qualitative properties. We also test the sensitivity of the strategies to implementation errors and misestimation of crucial parameters.

Scaling the parameters from the mean-field to the spatially explicit is not straightforward (41–43). The transmission rate in the mean-field case (β) is assumed to be the rate at which an individual susceptible site becomes infected by an infected neighbor, multiplied by the number of possible contacts in the neighborhood (four in this case) (32). In reality, the transmission rate depends on the number of neighboring sites, which is dynamically changing in time. For this reason, we cannot simply map the optimal strategies derived from the mean-field model onto the spatially explicit case. Instead, we concern ourselves with demonstrating that the qualitative nature of the optimal strategy in the mean-field case (Fig. 1a) is the same as in the contact process. We ran replicated simulations to find the

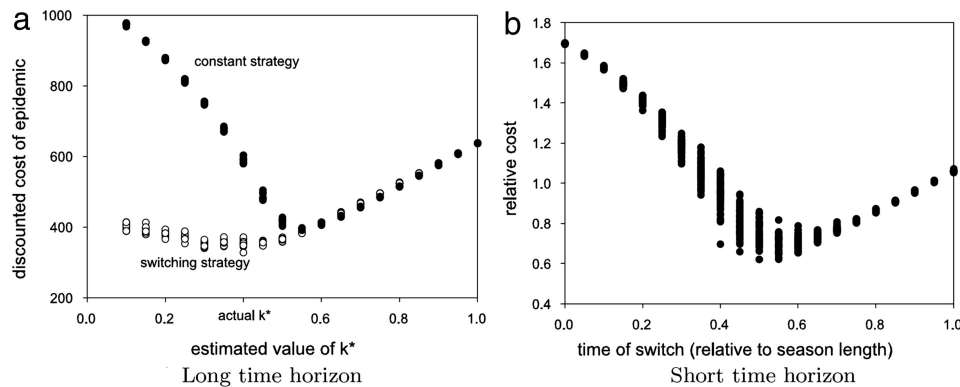


Fig. 3. Testing the robustness of the optimal strategies by using spatial simulations. (a) Total discounted cost of an epidemic vs. the estimated value of k^* . Two strategies are compared, a constant strategy where treatment remains at the estimated level of k^* throughout, and the two-phase strategy of switching from maximum treatment to k^* . Note that the epidemic runs with epidemiological parameters β and ϵ_i fixed at the correct values, so incorrect estimation of k^* results from not knowing these values. (b) Cost of an epidemic (relative to the most cost-effective constant strategy, which was determined numerically to be $\approx 55\%$ treatment) vs. the time of the switch (relative to the length of the season) from maximum to zero treatment. In both the short and long time frame cases, there were 50 realizations for each parameter set, with the default parameter values given in Table 1.

a simple representation of the spread of crop disease in a static landscape, it demonstrates the importance of economic considerations in epidemiological studies (28, 44) and allows us to extract some general principles that can then be tested in more realistic models for specific host–pathogen combinations.

The analyses give rise to recommendations for optimal strategies that differ from current conventions, where, for many diseases of annual and perennial crop plants, all infected fields are routinely treated. It is clear that simple analysis of epidemiological models would suggest that complete treatment of infected fields is not necessary if the objective is to prevent invasion of disease through the landscape or to prevent persistence (5, 45). Here the default for a system approximated by the contact process is to treat a proportion of fields, given for the mean-field approximation by k_c (Eq. 3), but doing so ignores the economic constraints associated with the cost of infection and the cost of treatment. Combining the two by using variational calculus (22) demonstrates that the epidemiological strategy is not necessarily optimal and leads to a set of alternative strategies, the outcome of which depends on the time frame of interest. The set of alternative strategies reflects, of course, the inclusion or otherwise of a discount rate (23) in the objective function, to be minimized. Here we have equated short-term regimes with a single season, for which we exclude the discount rate, comparing them with discounted longer-term regimes extending over multiple seasons. Note that it is the discount rate that separates the short- and long-term strategies, so the long time frame problem could just as easily be formulated as a finite time problem over more than one season (by replacing ∞ in Eq. 4 by large T). The results, in each case, depend on the relationships between the trade-offs in transmission and recovery rates for the epidemiological parameters and the ratio of costs for infection and treatment in the economic parameters (Figs. 1 and 2). Although both regimes can give rise to a singular solution, the criterion involving matching of the trade-offs in epidemiological with economic parameters is so exact for the short time horizon that we reject it as a plausible strategy. Doing so leaves only a bang-bang solution for the short time horizon provided $[\beta/(\epsilon_2 - \epsilon_1)] < (c/d)$. The singular solution holds, however, for the long time horizon (Fig. 1), with treatment of a fixed proportion ($k = k^*$) of infected fields, where $k^* < k_c$. In this case, the optimal strategy is a switch from an extreme level of treatment to a level comparable with the critical level required for pathogen extinction. Over a shorter time scale, it becomes nonoptimal to treat near criticality, and instead it is best to switch between extreme

treatments. This result is an intuitive one because the pathogen dynamics over a long time scale will eventually settle to equilibrium, whereas over short time scales, such as within a single growing season, the system will often remain in a transient state. Our model does not take account of the effect of interrupted transients associated with a long time scale for successive seasons of annual crops. Here, transients are routinely interrupted by harvesting, with successive epidemics being initiated by inoculum carried over from previous crops or by invasion from more distant sites for overwintering (4). The discontinuities remove the possibility of analytical results, but these systems are amenable to numerical investigation.

The relationship between the epidemiological and economic parameters determines a critical point at which a radically different strategy becomes optimal and is summarized for the long time frame case by a policy plot (Fig. 1a). The policy plot can be used to determine the nature of the optimal treatment regime, given prior knowledge of the epidemiological and economic parameters. The analytic results obtained from the mean-field formulation were tested on the spatially explicit contact-process model by using numerical simulation (Fig. 1b) and were found to be qualitatively identical but quantitatively different (because of the effects of spatial structure). It is important to note that the derived strategies are nonspatial. The advantage of nonspatial strategies is that they do not depend on a knowledge of the spatial location of disease. The strategies derived for the deterministic model outperformed the traditional constant policy and were shown to be robust against implementation errors. The robustness of the deterministic strategies was especially true in the case of long-term treatment, where misestimation of the interior level of treatment k^* had little effect on the overall cost of the epidemic.

Previous authors have found optimal singular solutions for epidemiological models with bifurcations. Both Sethi (21) and Goldman *et al.* (19) considered SIS models, whereas Barrett (28) analyzed a static economic model with bifurcation and a Nash equilibrium. The analyses presented here go beyond these models in analyzing the trade-off between the epidemiological and economic parameters for different time scales and in testing the robustness of the strategies in a spatially explicit scheme.

We made certain assumptions to simplify the optimization, such as a linear relationship between the cost and the amount of treatment and infection. If this relationship were not the case, it would lead to a more complicated optimal strategy that would dynamically change in time and hence would not be as practical

