CDDEP THE CENTER FOR Disease Dynamics, Economics & Policy
RESOURCES FOR THE FUTURE

Discussion Paper

March 2010 ■ RFF DP 10-15

Diversify or Focus? Spending to Combat Infectious Diseases When Budgets Are Tight

Soren Anderson, Ramanan Laxminarayan, and Stephen W. Salant

A center of
Resources for the Future

1616 P Street NW, Suite 600
Washington, DC 20036
p 202 328 5000 | f 202 939 3460
WWW.CDDEP.ORG

Diversify or Focus? Spending to Combat Infectious Diseases When Budgets Are Tight*

Soren Anderson†
Michigan State University

Ramanan Laxminarayan‡
Resources for the Future

Stephen W. Salant§
University of Michigan
Resources for the Future

February 2, 2010

Abstract

We consider a health authority seeking to allocate annual budgets optimally over time to minimize the discounted social cost of infection(s) evolving in a finite set of $R \geq 2$ groups. This optimization problem is challenging, since as is well known, the standard epidemiological model describing the spread of disease (SIS) contains a nonconvexity. Standard continuous-time optimal control is of little help, since a phase diagram is needed to address the nonconvexity and this diagram is $2R$ dimensional (a costate and state variable for each of the R groups). Standard discrete-time dynamic programming cannot be used either, since the minimized cost function is neither concave nor convex globally. We modify the standard dynamic programming algorithm and show how familiar, elementary arguments can be used to reach conclusions about the optimal policy with any finite number of groups. We show that under certain conditions it is optimal to focus the entire annual budget on one of the R groups at a time rather than divide it among several groups, as is often done in practice; faced with two identical groups whose only difference is their starting level of infection, it is optimal to focus on the group with fewer sick people. We also show that under certain conditions it remains optimal to focus on one group when faced with a wealth constraint instead of an annual budget.

*For helpful comments and suggestions, we thank seminar participants at CIREQ. We also thank Zachary Stangebye for valuable research assistance.

†sta@msu.edu or 517-355-0286
‡ramanan@rff.org or 202-328-5085
§ssalant@umich.edu or 734-223-7439

1 Introduction

Infectious disease remains an important cause of poor health in less developed countries, despite improvements in hygiene, sanitation, vaccination, and access to treatment (Lopez et al. 2006). Even in highly developed countries, diseases such as influenza and HIV/AIDS remain public health challenges. Although vaccines are available for some diseases, treating individuals who are already sick and trying to cure them is the only available intervention for many important diseases such as cholera, malaria, gonorrhea, and tuberculosis. Treatment, though not as effective as vaccination, is therefore an important tool in preventing the spread of infectious disease.

Unfortunately, limited budgets often constrain cost-effective treatment efforts. Unprecedented resources have been devoted to combating HIV, for example, yet the four million people in treatment represent less than 40 percent of those living with the disease. In Zambia, a country with one of the best-funded malaria control programs in sub-Saharan Africa, only 13 percent of children with malaria receive effective treatment. Such problems are pervasive. The World Health Organization (WHO), the Global Fund to Fight AIDS, TB and Malaria (GFATM), and individual ministries of health all operate under limited budgets. As a result, different countries or regions are often competing for the same funds. In the case of GFATM, for example, individual countries apply for money to fund disease control projects; some countries receive donated funds, while others do not.

When faced with such constraints and multiple infected populations, these agencies typically allocate funding in proportion to the number of people infected. The GFATM explicitly gives priority to low-income countries with high disease burden. This strategy seems equitable, but does it minimize the overall burden of disease? That is, to make the most of their limited budgets, should health authorities devote most treatment to groups with many infected people? Or should they focus on groups with many susceptible people? Or, as standard economic intuition might suggest, should they divide their budgets by equating the marginal impact of the last dollar of treatment spent on each group? Since the stated objective of these agencies is to reduce the burden of disease, the question of whether to diversify or focus is central to their missions.

In this paper we consider a health authority allocating treatment between two or more distinct groups to minimize the discounted social cost of infections over a finite time horizon. The infection in each group spreads according to the conventional SIS epidemiological model. The health authority can treat people at a constant marginal cost, but treatment each year is limited by a fixed annual budget. We show under tight budgets and other plausible assumptions that it is never optimal to divide the annual budget between the groups. Instead, the health authority should devote its entire budget in every period to just one group. Further, when there are two identical groups whose only difference is their starting level of infection, it is optimal to treat the sick in the group with the larger number of uninfected people. Since this group starts out healthier and gets all the treatment in the first period, it remains healthier in the subsequent period. Thus, as long as the budget remains insufficient to treat *every* infected individual, it is optimal to focus on the healthier group, period after period, to the complete neglect of sick people in the other group.

These results run counter to both conventional practice and standard economic intuition. They derive from the way an infection spreads, as described in the SIS model of disease. New infections arise from healthy people interacting with the sick. Thus, treating one sick person not only cures that single individual some percentage of the time but then also prevents healthy people from becoming infected at a later date. If there are many such healthy people, then spending the money required to treat one sick person prevents much disease. If many people are already sick, however, then treating one sick person prevents disease in fewer healthy people, and the treated individual herself is more likely to become sick again. The health authority in effect faces dynamic increasing returns to treatment in each group: the greater the number of healthy people, the more effective treating sick people in the group becomes. Thus, when presented with multiple infected groups and a limited budget, the health authority should take advantage of increasing returns by devoting its entire budget to a single group. Put differently, given the SIS dynamics, the health authority's cost-minimization problem is concave, leading to a corner solution in every period.

Determining how best to minimize the burden of infectious disease calls for a combination of epidemiological and economic insights, an approach taken in both the economics and the epidemiology literatures. Based on the pioneering work of Revelle (1967), Sanders (1971), and Sethi (1974), a more recent literature has emerged to clarify a number of important issues associated with this dynamic optimization problem (Goldman and Lightwood 2002; Rowthorn and Brown 2003; Gersovitz and Hammer 2004; Smith et al. 2005; Gersovitz and Hammer 2005; Herrmann and Gaudet 2009). None of these articles, however, describe the optimal treatment of multiple populations when the health authority has a limited budget.

Most of the literature minimizes the discounted sum of treatment costs plus the social costs of the infection. As always, the solution to such a "planning problem" is a valuable benchmark, since it identifies what is socially best. Often, however, health authorities in the real world are unable to achieve this first-best outcome. An authority may be charged, for example, with minimizing forgone production (or school attendance) due to illness but may lack the authority to tax or borrow. It then has no choice but to live within its annual budget. Indeed, governmental ministries of health may be prohibited by law from borrowing, as are entities such as the GFATM. In our base case, we assume that no one will lend to this health authority despite its promise to repay the loan out of its future annual budgets—perhaps because the health authority cannot precommit to repaying the loan in the future. To show that our results do not depend on this assumption, however, we also examine the less plausible case where the health authority can borrow against future budgets.[1]

The dynamic increasing returns to treatment inherent in the standard epidemiological model (SIS) make deriving the optimal treatment policy difficult, even for a single population. This nonconvexity in the planner's cost-minimization problem has haunted the literature from the outset. In an early paper, Sanders (1971) concludes that treatment should always be set to zero or the maximum possible level, but Sethi (1974), analyzing the same problem, concludes to the contrary that optimal treatment is always interior except in transitional phases at the beginning and end of the program. More recently, Gersovitz and Hammer (2004) recognize that they cannot prove analytically that the solutions to their necessary conditions are optimal, since, as they note, the standard sufficiency conditions fail in the

[1]This latter case involves the minimization of forgone production due to illness subject to any wealth constraint of the health authority. It therefore includes the solution of the planning problem as a special case; it also includes the case where the constraint is so tight that, at the constrained optimum, spending a dollar more in wealth would reduce the social cost of infection by more than a dollar.

presence of the nonconvexity (pp. 10 and 26). They instead rely on numerical simulations to argue that their solutions are likely optimal.

Goldman and Lightwood (2002) show how the nonconvexity inherent in the SIS dynamics can sometimes be handled. In the absence of diminishing returns to treatment, the Hamiltonian in their optimal control problem is linear in the control and, as they note, "comparisons must be made along all paths satisfying the necessary (or first order) conditions." They skillfully demonstrate this strategy in solving a planning problem involving a single population.

Although the same nonconvexity arises with multiple groups, Rowthorn et al. (2009) were unable to use the Goldman-Lightwood approach in their analysis of a health authority dividing its annual budget to treat the sick in two populations. Instead of the manageable two-dimensional phase diagram of Goldman and Lightwood (2002), their phase diagram would have been four dimensional. More generally, a phase diagram in $2R$ dimensions is needed: for each of the R groups, both the number who are sick and the associated co-state variable are changing over time. Like Gersovitz and Hammer (2004), Rowthorn et al. (2009) are, therefore, unable to prove analytically that the treatment policy they hypothesize to be optimal actually minimizes costs and are forced to rely on numerical simulations.[2]

The technical contribution of our paper is to show how to solve such problems. For expositional clarity, we first address the two-group problem formulated by Rowthorn et al. (2009). We then show how the same arguments easily extend to any finite number of groups. We use dynamic programming, but we modify the standard algorithm to circumvent a technical difficulty that would obscure understanding of the optimal policy rule: the "cost functions" generated in the backward recursion are not differentiable and are neither concave nor convex. Kinks and curvature problems arise whenever the health authority can treat every sick person in one group with budget left over.

In many circumstances, the prevalence of infection is so vast relative to available budgets that this troublesome situation would never arise. Even so, however, no qualitative conclusions about the optimal policy rule can be drawn using the standard dynamic programming algorithm. The standard algorithm would first establish properties of the cost function and policy rule that hold over the *entire* state space; only afterward would it use the initial condition and transition rule to determine the optimal trajectory through the subset of that space.

We therefore modify the standard algorithm to deduce results for the case of "tight budgets." In that case, starting from the initial conditions and following the SIS dynamics, it is *never* possible to treat all of the sick people in any group, even when devoting the entire budget to the same group in every period. In such cases, the infection pair always

[2]If the marginal cost of treatment increases sufficiently fast, then as (Gersovitz and Hammer 2004, p. 26) observe, it seems "intuitive" that the optimal treatment policy is interior. The requisite convexity of treatment costs in any given formulation, however, is not obvious. Assuming that the marginal cost of treatment is constant, therefore, seems to us an important base case. Moreover, it is commonplace in the literature. For example, Goldman and Lightwood (2002) showcase a linear cost function as a special case. Rowthorn et al. (2009) consider only linear costs. Sanders (1971) and Sethi (1974) are not directly comparable to these papers or ours, since they assume that costs are linear in the *fraction* of sick people receiving treatment. Finally, in our formulation where tight budgets might bind before any diminishing returns to treatment set in, the assumption of linear costs seems particularly appropriate.

lies within a given rectangular region. We establish that each cost function is strictly concave over this rectangular region; the irregularities in the cost function that undermine the standard algorithm arise elsewhere in the state space. Since no feasible treatment program can result in infection combinations outside this rectangular region, the curvature of the cost function outside the region is irrelevant. A similar argument may be useful in other dynamic programming problems for which qualitative conclusions about the optimal policy seem elusive.

The remainder of the paper is organized as follows. In section 2, we consider the problem of a health authority allocating funds over time to minimize the discounted sum of social costs of infection in two groups. Throughout this section, we assume that the health authority receives an annual budget and can neither borrow against it nor save funds for the future. In section 3, we show how our arguments generalize when the health authority allocates its budget across $R > 2$ groups. We then relax the constraint on borrowing and saving in section 4. Section 5 concludes the paper. Throughout, we assume that those infected in one group cannot transmit their disease to members of another group. Therefore, our analysis applies both to cases where every group faces the same disease, as well as to cases where the diseases in some (or all) of the groups are different.

2 The Health Authority's problem

2.1 Assumptions

Infectious diseases are spreading within two distinct groups. The spread of each infection is governed by the following difference equations:

$$I^i_{t+1} = (1 - \mu^i)I^i_t + \theta^i \frac{I^i_t}{N^i}(N^i - I^i_t) - \alpha^i F^i_t \tag{1}$$

where parameters $\mu^i, \theta^i, N^i, \alpha^i > 0$ for groups $i = A, B$. The population size of group i is fixed at N^i. The number of infected individuals (infecteds) in period t is I^i_t and the number of healthy individuals (susceptibles) is $N^i - I^i_t$. The number of individuals in group i treated in period t is F^i_t. We interpret μ^i as the fraction of infected individuals in group i who spontaneously recover (whether or not they were treated), $\theta^i I^i_t/N^i$ as the fraction of susceptibles that become infected (which is proportional to the fraction infected), and α^i as the fraction of those treated individuals whom the treatment cures (excluding those treated who would have recovered spontaneously so as to avoid double counting). Mathematically, each group's infection level is independent of the other's, which implies that the infections do not spread between the groups. In this case, different diseases can be afflicting the two groups.

We assume that the cost of treating each infected individual in group i is constant p^i and that the health authority has an annual budget in period t of M_t. The health authority can neither borrow against its budget nor save funds for the future. The budget constraint in period t is therefore

$$p^A F^A_t + p^B F^B_t \leq M_t, \tag{2}$$

for all $t = 1, ..., T$; we relax this assumption in Section 4. And we assume that the treatment is never used as a prophylactic, so only infected individuals receive it:

$$0 \leq F_t^i \leq I_t^i \tag{3}$$

for $i = A, B$ and all $t = 1, ..., T$.

To streamline the notation, let

$$\Gamma^i(I_t^i) \equiv (1 - \mu^i)I_t^i + \theta^i \frac{I_t^i}{N^i}(N^i - I_t^i) \tag{4}$$

denote the number of infections in group i in period $t+1$ if there were I_t^i infecteds in period t and none received treatment. Then the spread of the infection in group i can be re-expressed as

$$I_{t+1}^i = \Gamma^i(I_t^i) - \alpha^i F_t^i \tag{5}$$

for $i = A, B$. For most of the paper we suppress further mention of the specific functional form in (1) above. Thus, all of our results will hold for the generic functional form in (5), subject to a handful of regularity assumptions that we detail immediately below.

We impose several restrictions on the infection dynamics. First, we assume that $\Gamma^i(I_t^i) - \alpha^i I_t^i$ is strictly increasing in I_t^i for all $I_t^i \in [0, N^i]$ for $i = A, B$. That is, if every infected individual received treatment, the more infected people there are in the current period, the more there would be in the next period. As we show below, this assumption guarantees that the cost function is increasing, which implies that it is always optimal to exhaust the budget.[3] It follows from this assumption that $\Gamma^i(I_t^i)$ is also increasing for all $I_t^i \in [0, N^i]$ for $i = A, B$.

Second, we assume that $\Gamma^i(I_t^i)$ is strictly concave for all $I_t^i \in [0, N^i]$ for $i = A, B$. That $\Gamma(\cdot)$ is concave when we impose our specific functional form is clear from the second derivative $-\frac{2\theta^i}{N^i} < 0$ and requires no further assumptions in that special case.

Third, we assume that $\Gamma^i(0) = 0$ and that $\Gamma^i(N^i) \leq N^i$ for $i = A, B$, which, given our other assumptions, implies that infection levels in each group are always nonnegative and never exceed either group's population:

$$0 \leq I_t^i \leq N^i \tag{6}$$

for $i = A, B$ and all $t = 1, ..., T$.[4]

Finally, we make the following simplifying assumption: even if the health authority devoted its entire budget in every period to the same group (either A or B), it would never have enough funds to treat all the infected individuals in that group. Hence, no matter how the health authority behaves,

[3]Given the specific functional form we present above, this assumption reduces to a restriction on the first derivative evaluated at $I_t^i = N^i$:

$$1 - \mu^i - \theta^i - \alpha^i > 0,$$

for $i = A, B$.

[4]It is clear that $\Gamma^i(0) = 0$ for our specific functional form. The assumption that $\Gamma^i(N^i) \leq N^i$ holds as long as $\mu^i \in [0, 1]$ for $i = A, B$: given an infection level of $I_t^i \in [0, N^i]$ in period t, the minimum number of infecteds in period $t + 1$ would be $\Gamma^i(0) - \alpha^i \cdot 0 = 0$, while the maximum would be $\Gamma^i(N^i) = (1 - \mu^i)N^i \in [0, N^i]$.

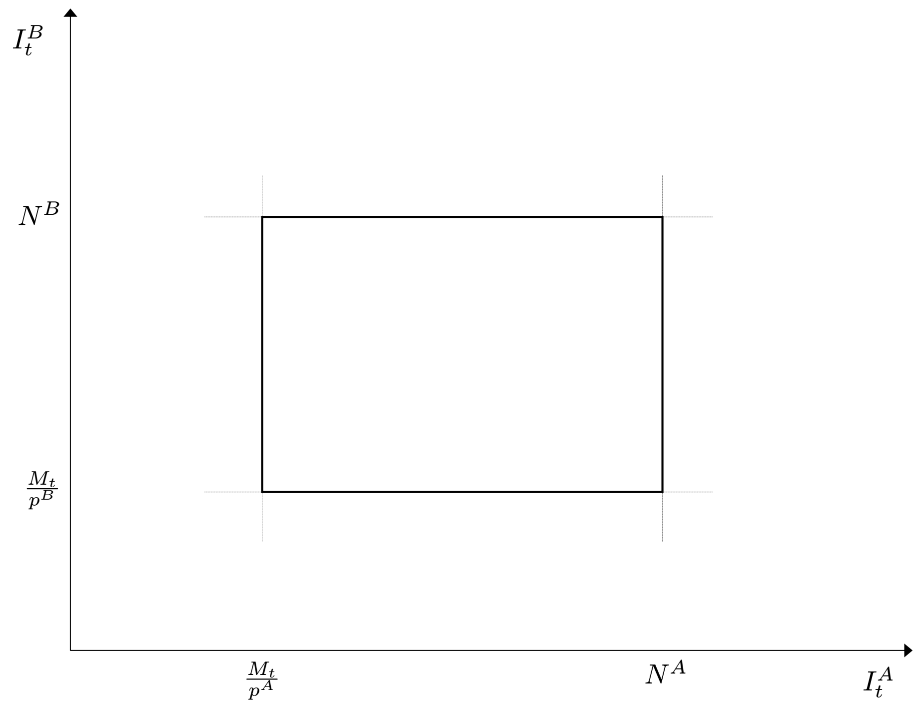

Figure 1: Period t's rectangle

Note: Assumption 1 is that it is never possible to treat all the sick people in either group. Thus, infection pairs in period t must lie within the rectangular region identified in the figure. Given assumption 1, we are able to establish concavity of the cost function over this rectangular region.

Assumption 1. $N^i \geq I_t^i \geq \frac{M_t}{p^i}$ *for* $i = A, B$ *for all* $t = 1, 2, ..., T$.

It follows that, in period t, the infection pair must lie in a rectangular region of height $N^B - \frac{M_t}{p^B}$ and length $N^A - \frac{M_t}{p^A}$. We refer to this as "period t's rectangle." See Figure 1. Assumption 1 is not necessary in some cases for our result but seems realistic and simplifies the analysis.

2.2 Cost-minimization problem

The health authority chooses how many individuals to treat in each group in every period $t = 1, ..., T$, subject to constraints (2) and (3) above, to minimize the discounted sum of the social costs of infection in the two groups:

$$\sum_{t=1}^{T} \delta^{t-1}[s^A I_t^A + s^B I_t^B],$$

where T indexes the final period, s^i denotes the social cost per infected person in group i, and $\delta \in (0, 1)$ denotes the discount factor. We assume that infection levels initially are $\bar{I}_1^A$

and $\bar{I}_1^B$ and thereafter follow the relevant difference equation in (1) (or equation (5), the case of a generic functional form).

It is convenient to express the health authority's problem recursively. Denote the minimized cost of entering period $t+1$ with infection pair (I_{t+1}^A, I_{t+1}^B) and proceeding optimally thereafter as $C_{t+1}(I_{t+1}^A, I_{t+1}^B)$.[5] The cost function for period t therefore is given by

$$C_t(I_t^A, I_t^B) = \min_{F_t^A, F_t^B} s^A I_t^A + s^B I_t^B + \delta C_{t+1}(\Gamma^A(I_t^A) - \alpha^A F_t^A, \Gamma^B(I_t^B) - \alpha^B F_t^B)$$

subject to $p^A F_t^A + p^B F_t^B \leq M_t$ and $F_t^A, F_t^B \geq 0$.

2.3 Three-period problem

To build intuition, we begin by determining what the health authority should do in the last three periods ($T-2, T-1,$ and T). Since spending at T has no effect on social costs until after the end of the horizon, it is optimal to spend nothing then ($F_T^i = 0$) and $C_T(I_T^A, I_T^B) = s^A I_T^A + s^B I_T^B$. As for the optimal decision at $T-1$, the social cost of infection is a linear, strictly decreasing function of the number of individuals (F_{T-1}^A, F_{T-1}^B) treated in each group. Hence, it will be optimal to spend the entire budget on only one group.

More formally, the health authority wants to allocate its budget at $T-1$ to minimize the discounted social cost of infection from that period onward:

$$\begin{aligned} C_{T-1}(I_{T-1}^A, I_{T-1}^B) = \min_{F_{T-1}^A, F_{T-1}^B} \; & s^A I_{T-1}^A + s^B I_{T-1}^B \\ & + \delta[s^A \Gamma^A(I_{T-1}^A) - s^A \alpha^A F_{T-1}^A + s^B \Gamma^B(I_{T-1}^B) - s^B \alpha^B F_{T-1}^B] \end{aligned} \tag{7}$$

subject to $F_{T-1}^A, F_{T-1}^B \geq 0$ and $p^A F_{T-1}^A + p^B F_{T-1}^B \leq M_{T-1}$.

The constraint set is a familiar budget triangle, with boundary slope $-\frac{p^A}{p^B}$. The slope indicates the number of additional infected individuals in group B who can be treated using the money saved by not treating one individual in group A.[6] The preference direction is also conventional (northeast), since treating more individuals in either group reduces the discounted sum of social costs of infection in the two groups. The indifference curves are downward-sloping lines with $MRS = -\frac{\alpha^A s^A}{\alpha^B s^B} < 0$. So the health authority should spend its

[5] We have suppressed the dependence of minimized costs on future budget levels and various fixed parameters to simplify the notation.

[6] Assumption 1 ensures that *all treatment pairs* in the budget triangle are feasible. If Assumption 1 were violated, there could be money left over after treating every infected individual in one group. In that case, treatment pairs in one (or both) corners of the triangle would be deemed inadmissible, since the health authority does not treat susceptibles prophylactically. That is, the constraint set is the intersection of *three* constraints: the budget set *and* $F_t^i \leq I_t^i$ for $i = A, B$. Assumption 1 guarantees that these additional constraints never bind and can be disregarded. If Assumption 1 is violated, the optimum might occur where $F_t^i = I_t^i < \frac{M_t}{p^i}$. In that case, funds remaining would be spent on the other group. This not only increases the number of cases that must be considered but also results in a minimized cost function that is not concave everywhere. That the health authority never serves both groups (unless it has treated every infected individual in one group) can still be established in the only case we have examined (the symmetric case). But concavity arguments no longer work, and the proof instead relies on a property of the cost function that seems of limited applicability to other problems. The proof for the symmetric case in the absence of Assumption 1 is available upon request.

entire budget on group A if $\frac{\alpha^A s^A}{p^A} > \frac{\alpha^B s^B}{p^B}$ and on group B if $\frac{\alpha^A s^A}{p^A} < \frac{\alpha^B s^B}{p^B}$. If neither inequality holds, then the health authority is indifferent between spending on infected individuals in either group and any combination of treatment that exhausts the budget is optimal. Intuitively, by treating one less individual in group A, the health authority can treat $\frac{p^A}{p^B}$ more individuals in group B. If one less individual is treated in group A, discounted social costs there would increase by $\alpha^A s^A$; if $\frac{p^A}{p^B}$ more individuals are treated in group B, social costs there would decrease by $\alpha^B s^B \frac{p^A}{p^B}$. This reallocation is beneficial if the result is a *net* reduction in discounted social costs from that period onward ($\alpha^B s^B \frac{p^A}{p^B} - \alpha^A s^A > 0$); if, instead, it *raises* social costs, then the arbitrage should be reversed. Reallocating toward group A is optimal if (1) the price of treatment (p^A) in group A is sufficiently low, (2) the effectiveness of the treatment (α^A) in group A is sufficiently high, or (3) the social cost of infection (s^A) in group A is sufficiently high.

The decision in period $T-1$ of how many individuals in each group to treat is straightforward because the only cost consequences occur in period T. When the decision is made in some prior period, however, the health authority should take into account the discounted cost consequences from the next period onward of reallocating the current budget. To deduce these consequences, the authority needs to compute the sequence of minimized cost functions.

We can compute the minimized cost function in period $T-1$ by substituting the optimal decision rule (F^A_{T-1}, F^B_{T-1}) into the objective function in equation (13):

$$\begin{aligned} C_{T-1}(I^A_{T-1}, I^B_{T-1}) &= s^A I^A_{T-1} + s^B I^B_{T-1} \\ &+\delta \left[s^A \Gamma^A(I^A_{T-1}) + s^B \Gamma^B(I^B_{T-1}) - M_{T-1} \max\left(\frac{\alpha^A s^A}{p^A}, \frac{\alpha^B s^B}{p^B} \right) \right]. \end{aligned} \tag{8}$$

The cost function $C_{T-1}(I^A_{T-1}, I^B_{T-1})$ is the sum of continuous functions and is therefore continuous. Since $\Gamma^i(\cdot)$ is strictly increasing in its single argument (for $i = A, B$), $C_{T-1}(I^A_{T-1}, I^B_{T-1})$ is strictly increasing in its two arguments. Moreover, since $\Gamma^i(\cdot)$ is strictly concave in its single argument and the right-hand side of (8) is additively separable, $C_{T-1}(I^A_{T-1}, I^B_{t-1})$ is strictly jointly concave (henceforth, simply "strictly concave") in (I^A_{T-1}, I^B_{T-1}). As we will show, prior minimized cost functions inherit these two properties; each is strictly increasing and strictly concave.

The health authority allocates its budget at $T-2$ to minimize the discounted social cost of infection from that period onward:

$$\begin{aligned} C_{T-2}(I^A_{T-2}, I^B_{T-2}) = &\min_{F^A_{T-2}, F^B_{T-2}} s^A I^A_{T-2} + s^B I^B_{T-2} \\ &+\delta C_{T-1}(\Gamma^A(I^A_{T-2}) - \alpha^A F^A_{T-2}, \Gamma^B(I^B_{T-2}) - \alpha^B F^B_{T-2}) \end{aligned} \tag{9}$$

subject to $F^A_{T-2}, F^B_{T-2} \geq 0$ and $p^A F^A_{T-2} + p^B F^B_{T-2} \leq M_{T-2}$. Since the constraint set is compact and the minimand is continuous, an optimal allocation rule for the budget at $T-2$ exists (Weierstrass Theorem) and the minimum value function $C_{T-2}(\cdot,\cdot)$ is itself continuous (Berge's Theorem).[7] Admissible allocations of the period $T-2$ budget between

[7]Proceeding inductively, it is straightforward to establish that an optimal budget allocation rule exists at each stage.

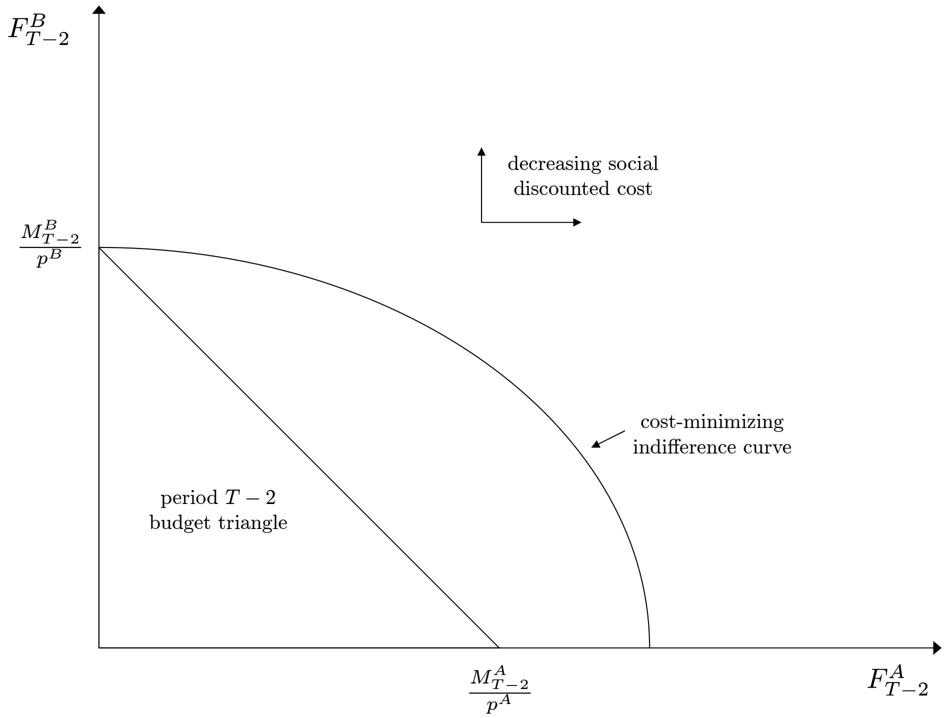

Figure 2: Allocation of the health budget in period $T-2$ and its consequences for social cost

the two groups lie in a right-triangle with hypotenuse of slope -$\frac{p^A}{p^B}$. See Figure 2. The locus of expenditures across the two groups that results in a given discounted social cost from the current period onward is an "indifference curve." The indifference curve is downward sloping, since spending more on either group would lower the discounted social cost. Budget allocations to the "southwest" of the indifference curve (including the origin, where nothing is spent on the sick of either group) result in strictly higher social cost while budget allocations to the "northeast" result in strictly lower social cost. Since $C_{T-1}(I^A_{T-1}, I^B_{T-1})$ is strictly concave in the infection pair, it is strictly concave in the pair of health expenditures and must be strictly quasi-concave in them as well. Hence, budget allocations resulting in a strictly higher discounted social cost must form a strictly convex set, and the optimal allocation of the health budget requires spending the entire health budget at $T-2$ on a single group.[8]

[8] The marginal rate of substitution between F^A_{T-2} and F^B_{T-2} is:

$$\frac{dF^B_{T-2}}{dF^A_{T-2}} = -\frac{(\delta\alpha^A + \delta^2\alpha^A\Gamma^{A\prime}(\Gamma^A(I^A_{T-2}) - \alpha^A F^A_{T-2}))s^A}{(\delta\alpha^B + \delta^2\alpha^B\Gamma^{B\prime}(\Gamma^B(I^B_{T-2}) - \alpha^B F^B_{T-2}))s^B}, \tag{10}$$

which is strictly negative since $\Gamma^i(\cdot)$ is strictly increasing for $i = A, B$. Since $\Gamma^i(\cdot)$ is also strictly concave, the magnitude of the MRS strictly *increases* as one moves downward along any indifference curve (as F^A_{T-2} increases and F^B_{T-2} decreases). That is, the indifference curves are strictly concave.

2.4 Problem of any finite length

With any finite number of periods, the health authority chooses $F_t^A, F_t^B \geq 0$ to minimize

$$s^A I_t^A + s^B I_t^B + \delta C_{t+1}\left(\Gamma^A(I_t^A) - \alpha^A F_t^A, \Gamma^B(I_t^B) - \alpha^B F_t^B\right) \tag{11}$$

subject to the budget constraint $p^A F_t^A + p^B F_t^B \leq M_t$. We will show that the cost function in every period ($t = 1, \ldots T-1$) is both strictly increasing and strictly concave in its inherited infection levels. This in turn implies that in every period the health authority will devote its entire budget to either one group or the other.

We begin by showing that the cost function is strictly increasing in every period.

Theorem 1. *The cost function $C_t(I_t^A, I_t^B)$ is strictly increasing for any $I_t^i \in [\frac{M_t}{p^i}, N^i]$ where $i = A, B,$ and $t = 1, 2, ..., T$.*

Proof. First recall that the cost function at $T-1$, $C_{T-1}(I_{T-1}^A, I_{T-1}^B)$, is strictly increasing for any infection levels (I_{T-1}^A, I_{T-1}^B) in the period $T-1$ rectangle. Now assume inductively that in any period $t < T-1$ the cost function starting in the subsequent period $C_{t+1}(I_{t+1}^A, I_{t+1}^B)$ is strictly increasing for any infection levels (I_{t+1}^A, I_{t+1}^B) in the period $t+1$ rectangle. We will show that the cost function starting in period t given by $C_t(I_t^A, I_t^B)$ is strictly increasing for any infection levels (I_t^A, I_t^B) in the period t rectangle.

Let

$$\Omega(I_t^A, I_t^B, F_t^A, F_t^B) \equiv s^A I_t^A + s^B I_t^B + \delta C_{t+1}\left(\Gamma^A(I_t^A) - \alpha^A F_t^A, \Gamma^B(I_t^B) - \alpha^B F_t^B\right) \tag{12}$$

denote the objective function that the health authority minimizes. Note that by hypothesis $C_{t+1}(I_{t+1}^A, I_{t+1}^B)$ is strictly increasing for (I_{t+1}^A, I_{t+1}^B) in the period $t+1$ rectangle and that $\Gamma^i(I_t^i)$ is strictly increasing in I_t^i for $i = A, B$. Thus, $\Omega(I_t^A, I_t^B, F_t^A, F_t^B)$ is strictly increasing for any infection pair (I_t^A, I_t^B) in the period t rectangle. Let $(\hat{F}_t^A, \hat{F}_t^B)$ be the cost-minimizing treatment levels given initial infection levels $(\hat{I}_t^A, \hat{I}_t^B)$ in the period t rectangle. Consider any smaller infection pair (I_t^A, I_t^B) in the period t rectangle such that $\hat{I}_t^A > I_t^A$ and $\hat{I}_t^B = I_t^B$, so that the infection pair is strictly smaller in the direction of I_t^A. Then the minimized discounted social cost from period t onward of entering period t with a strictly smaller infection pair is strictly smaller. That is,

$$\begin{aligned} C_t(\hat{I}_t^A, \hat{I}_t^B) &= \Omega(\hat{I}_t^A, \hat{I}_t^B, \hat{F}_t^A, \hat{F}_t^B) \\ &> \Omega(I_t^A, I_t^B, \hat{F}_t^A, \hat{F}_t^B) \\ &\geq C_t(I_t^A, I_t^B), \end{aligned}$$

where equality in the first line follows from the definition of the cost function, the inequality in the second line follows from the fact that the minimand is an increasing function of initial infection levels, and the inequality in the third line follows from cost minimization. So the cost function is strictly increasing for any $I_t^A \in [\frac{M_t}{p^A}, N^A]$. A symmetric argument establishes that it is also strictly increasing for any $I_t^B \in [\frac{M_t}{p^B}, N^B]$.[9] □

[9] We could show that the cost function is strictly increasing over a wider range of infection levels, but the proof of our key result below relies on concavity, which we are only able to establish over the period t rectangle.

Next we show that the cost function in every period is strictly concave over the rectangular region identified in assumption 1.

Theorem 2. *The cost function $C_t(I_t^A, I_t^B)$ is strictly concave for any $I_t^i \in [\frac{M_t}{p^i}, N^i]$ where $i = A, B,$ and $t = 1, 2, ..., T-1$.*

Proof. Recall that $C_{T-1}(I_{T-1}^A, I_{T-1}^B)$, defined in equation (8), is strictly concave for any (I_{T-1}^A, I_{T-1}^B) in the period $T-1$ rectangle. Now assume inductively that in any period $t < T-1$ the cost function starting in the subsequent period $C_{t+1}(I_{t+1}^A, I_{t+1}^B)$ is strictly concave for any (I_{t+1}^A, I_{t+1}^B) in the period $t+1$ rectangle. We will show that the cost function starting in period t given by $C_t(I_t^A, I_t^B)$ is strictly concave for any (I_t^A, I_t^B) in the period t rectangle.

As before, let

$$\Omega(I_t^A, I_t^B, F_t^A, F_t^B) \equiv s^A I_t^A + s^B I_t^B + \delta C_{t+1}\left(\Gamma^A(I_t^A) - \alpha^A F_t^A, \Gamma^B(I_t^B) - \alpha^B F_t^B\right)$$

be the function that the health authority minimizes. Note that if $C_{t+1}(I_{t+1}^A, I_{t+1}^B)$ is strictly concave in (I_{t+1}^A, I_{t+1}^B) for any infection pair in the period $t+1$ rectangle, then $\Omega(I_t^A, I_t^B, F_t^A, F_t^B)$ is strictly concave for any (I_t^A, I_t^B) in the period t rectangle, since $\Gamma^i(I_t^i)$ is strictly concave in I_t^i for $i = A, B$.[10]

By the definition of strict concavity,

$$\Omega(\lambda I_t^A + (1-\lambda)\hat{I}_t^A, \lambda I_t^B + (1-\lambda)\hat{I}_t^B, F_t^A, F_t^B) > \lambda\Omega(I_t^A, I_t^B, F_t^A, F_t^B) + (1-\lambda)\Omega(\hat{I}_t^A, \hat{I}_t^B, F_t^A, F_t^B)$$

for any two distinct feasible pairs of starting infection levels (I_t^A, I_t^B) and $(\hat{I}_t^{\,A}, \hat{I}_t^{\,B})$ in the period t rectangle, any constant $\lambda \in (0,1)$, and any affordable treatment level (F_t^A, F_t^B).

Now let

$$(F_t^{A\lambda}, F_t^{B\lambda}) = \underset{(F_t^A, F_t^B)}{\arg\min}\, \Omega(\lambda I_t^A + (1-\lambda)\hat{I}_t^A, \lambda I_t^B + (1-\lambda)\hat{I}_t^B, F_t^A, F_t^B)$$

be the cost-minimizing treatment starting from initial infection levels $(\lambda I_t^A + (1-\lambda)\hat{I}_t^A, \lambda I_t^B + (1-\lambda)\hat{I}_t^B)$. Then

$$\begin{aligned} C_t(\lambda I_t^A + (1-\lambda)\hat{I}_t^A, \lambda I_t^B + (1-\lambda)\hat{I}_t^B) &= \Omega(\lambda I_t^A + (1-\lambda)\hat{I}_t^A, \lambda I_t^B + (1-\lambda)\hat{I}_t^B, F_t^{A\lambda}, F_t^{B\lambda}) \\ &> \lambda\Omega(I_t^A, I_t^B, F_t^{A\lambda}, F_t^{B\lambda}) + (1-\lambda)\Omega(\hat{I}_t^A, \hat{I}_t^B, F_t^{A\lambda}, F_t^{B\lambda}) \\ &\geq \lambda C_t(I_t^A, I_t^B) + (1-\lambda) C_t(\hat{I}_t^A, \hat{I}_t^B), \end{aligned}$$

[10] Observe that

$$\begin{aligned} &C_{t+1}(\Gamma^A(\lambda I_t^A + (1-\lambda)\hat{I}_t^A) - \alpha^A F_t^A, \Gamma^B(\lambda I_t^B + (1-\lambda)\hat{I}_t^B) - \alpha^B F_t^B) > \\ &C_{t+1}(\lambda(\Gamma^A(I_t^A) - \alpha^A F_t^A) + (1-\lambda)(\Gamma^A(\hat{I}_t^A) - \alpha^A F_t^A), \lambda(\Gamma^B(I_t^B) - \alpha^B F_t^B) + (1-\lambda)(\Gamma^B(\hat{I}_t^B) - \alpha^B F_t^B) > \\ &\lambda C_{t+1}(\Gamma^A(I_t^A) - \alpha^A F_t^A, \Gamma^B(I_t^B) - \alpha^B F_t^B) + (1-\lambda) C_{t+1}(\Gamma^A(\hat{I}_t^A) - \alpha^A F_t^A, \Gamma^B(\hat{I}_t^B) - \alpha^B F_t^B) \end{aligned}$$

for any two sets of distinct infection pairs (I_t^A, I_t^B) and $(\hat{I}_t^A, \hat{I}_t^B)$ in the period t rectangle and any $\lambda \in (0,1)$. The first inequality follows from $\Gamma^i(\cdot)$ strictly concave for $i = A, B$ and $C_{t+1}(I_{t+1}^A, I_{t+1}^B)$ strictly increasing (note the substitution $F_t^i = \lambda F_t^i + (1-\lambda) F_t^i$), while the second inequality follows from $C_{t+1}(I_{t+1}^A, I_{t+1}^B)$ strictly concave in its arguments. Thus, the first and last lines together imply that $C_{t+1}(\Gamma^A(I_t^A) - \alpha^A F_t^A, \Gamma^B(I_t^B) - \alpha^B F_t^B)$ is strictly concave in (I_t^A, I_t^B). Thus, $\Omega(I_t^A, I_t^B, F_t^A, F_t^B)$ is also strictly concave in (I_t^A, I_t^B), since it simply adds a linear combination of I_t^A and I_t^B to $C_{t+1}(\Gamma^A(I_t^A) - \alpha^A F_t^A, \Gamma^B(I_t^B) - \alpha^B F_t^B)$.

where the equality in the first line follows from the definition of the cost function, the inequality in the second line follows from strict concavity, and the inequality in the third line follows from cost minimization. Therefore, $C_t(I_t^A, I_t^B)$ is strictly concave in (I_t^A, I_t^B).[11] $\square$

We now pull together the several results above to prove the key qualitative result of our analysis that in every period the health authority will focus its entire budget on one group or the other.

Theorem 3. *Under assumption 1, it is optimal either for* $F_t^A = \frac{M_t}{p^A}$ *and* $F_t^B = 0$ *or* $F_t^A = 0$ *and* $F_t^B = \frac{M_t}{p^B}$ *in every period* $t = 1, 2, ..., T-1$.

Proof. Because $C_{t+1}(I_{t+1}^A, I_{t+1}^B)$ is strictly increasing for any (I_{t+1}^A, I_{t+1}^B) in the period $t+1$ rectangle, the health authority's minimand in equation (14) is strictly decreasing in both F_t^A and F_t^B (by theorem 1) and the health authority should spend its entire budget in period t. We already showed that spending the entire budget on one group is optimal in period $T-1$. Moving to earlier periods, since the cost function starting in period $t+1$ is strictly concave for any (I_{t+1}^A, I_{t+1}^B) in the period $t+1$ rectangle (by theorem 2), it is also strictly concave in F_t^A and F_t^B for every budget allocation that is affordable in period t. This follows because $\Gamma^i(I_t^i) - \alpha^i F_t^i$ is linear in F_t^i for $i = A, B$. Consequently the indifference curves are strictly concave everywhere in the budget set. Consequently, spending the entire budget on one group is always optimal. $\square$

Thus far, we have established that it is optimal for the health authority to allocate each period's budget entirely to one group or the other. This result follows from the concavity of the cost function over the relevant domain. To determine *which* group will receive treatment, however, we must impose further assumptions. If we assume that the groups share identical infection dynamics, treatment price, and social cost, such that the only difference is the starting level of infection, it turns out that it is optimal to devote the entire budget to the group with the *lower* level of infection, and this same group will receive treatment in every period. The following theorem and proof establish this result.

Theorem 4. *Under assumption 1, if group A and group B have identical parameters (i.e., if* $\Gamma^A(\cdot) = \Gamma^B(\cdot) = \Gamma(\cdot)$, $\alpha^A = \alpha^B = \alpha$, $p^A = p^B = p$, *and* $s^A = s^B = s$*), then it is optimal to treat the group in which the current rate of infection is lowest in every period* $t = 1, ..., T-1$, *and the same group will receive treatment in every period* $t = 1, ..., T-1$.[12]

[11]Note that assumption 1 is critical here as it implies that the constraint set is the same for all pairs of inherited infection levels and their convex combinations: $p^A F_t^A + p^B F_t^B \leq M_t$. Without this assumption, there would be no guarantee that $(F_t^{A\lambda}, F_t^{B\lambda})$ is feasible from both (I_t^A, I_t^B) and $(\hat{I}_t^A, \hat{I}_t^B)$, in which case the second and third lines above would make no sense. For example, suppose it were the case that $I_t^A, I_t^B > 0$ and $\hat{I}_t^A, \hat{I}_t^B = 0$. In this scenario, $(F^{A\lambda}, F^{B\lambda})$ is not even feasible from $(\hat{I}_t^A, \hat{I}_t^B)$, given the constraints that $0 \leq F_t^i \leq I_t^i$ for $i = A, B$, and the proof breaks down. In addition, without assumption 1, the cost function in period $T-1$ is not globally concave over all relevant starting infection pairs, as we showed above, so the proof by induction never gets off the ground. For a further discussion of assumption 1, see the Appendix.

[12]It turns out that this result holds in the absence of assumption 1, with the modification that treatment is allocated only to the group with higher levels of infection after every sick person in the other group has already been treated. As we note above, the proof relies on a nonstandard property of the cost function, which is specific to the case where group A and group B have identical parameters and therefore of limited applicability to other problems. This proof is available upon request.

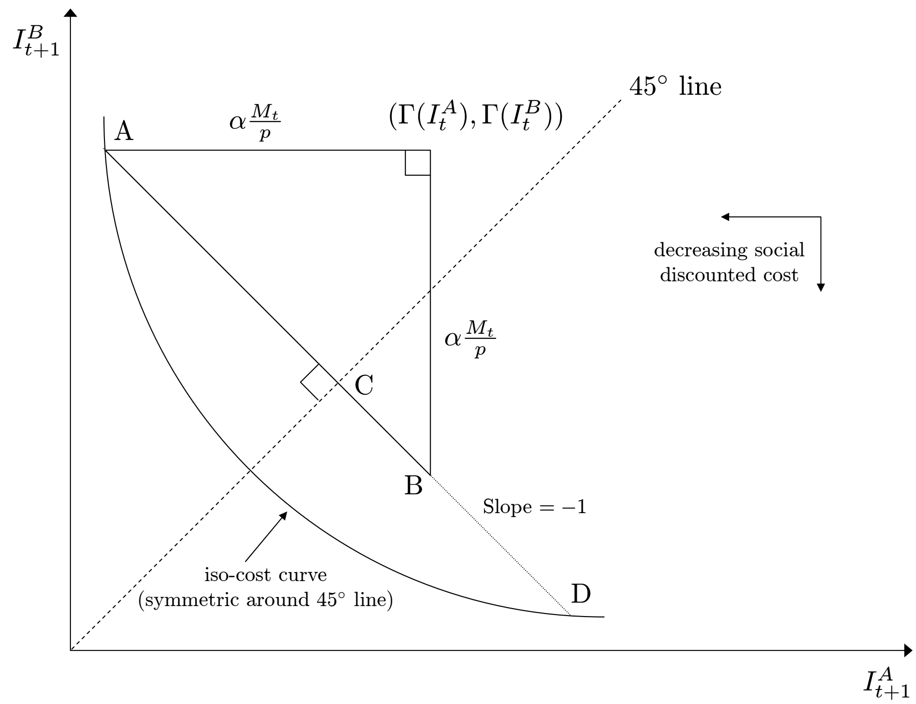

Figure 3: When other parameters are identical, treating the group with fewer sick people is optimal

Note: See text for details.

Proof. Suppose, without loss of generality, that there are strictly fewer sick people in group A; that is, $I^A_t < I^B_t$. Then, because $\Gamma^i(\cdot)$ is increasing and identical for $i = A, B$, we know that, if no one receives treatment, infection will remain lower in group A in the next period; that is, $\Gamma^A(I^A_t) < \Gamma^B(I^B_t)$. Therefore, if no one receives treatment, next period's infections will be located at infection pair $(\Gamma^A(I^A_t), \Gamma^B(I^B_t))$, to the left of the 45-degree line. See figure 3.

We have shown that it is optimal to spend the entire budget on group A or on group B. If the budget is spent on group A, then next period's infection pair will lie at $(\Gamma(I^A_t) - \alpha\frac{M_t}{p}, \Gamma(I^B_t))$. This pair is labeled point A in the figure. If the budget is spent on group B, then next period's infection pair will lie at $(\Gamma(I^A_t), \Gamma(I^B_t) - \alpha\frac{M_t}{p})$, which is labeled point B in the figure. Since spending in period t affects infections only in period $t+1$, our problem reduces to determining which of the following is lower: the period $t+1$ cost function evaluated at point A (given by C_{t+1}(point A)) or the cost function evaluated at point B (given by C_{t+1}(point B)).

Reflect point A across the 45-degree line and label this new point D. Now draw a line through points A and D. This line has slope -1 and crosses the 45-degree line at point C. Point B must lie on segment AD. Why? Because points A and B are equidistant to point $(\Gamma^A(I^A_t), \Gamma^B(I^B_t))$, which means a line drawn through points A and B also has slope -1.

Finally, observe that the cost function evaluated at point A must equal the function evaluated at point D; that is, $C_{t+1}(\text{point A}) = C_{t+1}(\text{point D})$, since the cost function is symmetric around the 45-degree line (by identical group parameters), and since point D is simply point A's reflection across the line. Thus, it must be that $C_{t+1}(\text{point B}) > C_{t+1}(\text{point A})$, since point B lies on segment AD, since costs at point A equal costs at point D, and since the cost function is strictly concave. We have drawn point B below the 45-degree line, although the same logic obviously holds if point B is on segment AD but above the 45-degree line.

Thus, it is optimal to spend the entire budget treating sick people in group A. By symmetry, it would be optimal to spend the budget on group B if its rate of infection were strictly lower. So it is optimal to treat the group with fewer sick people in every period. And obviously, if both groups had the same rate of infection initially, then the health authority would be indifferent between treating the sick in group A or the sick in group B in that initial period.

It remains to show that the same group will receive treatment in every period. Suppose that $I_t^A \leq I_t^B$, such that in period t it is optimal to treat group A. Then there will be strictly fewer infections in group A in period $t+1$ because $\Gamma(I_t^A) - \alpha M_t/p < \Gamma(I_t^B)$, by $\Gamma(\cdot)$ increasing and assuming $M_t > 0$. Since group A will have fewer infections in period $t+1$, we have shown that it will again receive treatment in that period, and it will therefore have fewer infections in period $t+2$, and so on. By symmetry, group B would receive treatment in every period if it received treatment first. □

3 Generalization to $R > 2$ groups

Suppose there are $R > 2$ groups. We will again assume that the budget is too small to treat all the infected people in any group in any period, so the $R > 2$ equivalent of assumption 1 still holds: $N^i \geq I_t^i \geq \frac{M_t}{p^i}$ for $i = 1, \ldots, R$ for all $t = 1, 2, ..., T$. Now the period t rectangle is an R-dimensional "hyper rectangle" with dimensions $N^i - \frac{M_t}{p^i}$ for $i = 1, \ldots, R$.

Now reconsider the optimal decision at $T-1$. Since the social cost of infection is a linear, strictly decreasing function of the number of infected individuals $(F_{T-1}^i, \ldots, F_{T-1}^R)$ treated in each group, it will be optimal to spend the entire budget at $T-1$ on only one group.

More formally,

$$C_{T-1}(I_{T-1}^1, \ldots, I_{T-1}^R) = \min_{F_{T-1}^1, \ldots, F_{T-1}^R} \sum_{i=1}^R s^i I_{T-1}^i + \delta \sum_{i=1}^R s^i [\Gamma^i(I_{T-1}^i) - \alpha^i F_{T-1}^i] \qquad (13)$$

subject to $F_{T-1}^i \geq 0$ and $\sum_{i=1}^R p^i F_{T-1}^i \leq M_{T-1}$. Since unspent budget has no value and spending on the sick people in any group strictly reduces infection and hence social cost ($\alpha^i > 0$), the entire budget will be spent. Since spending on group i lowers the social cost of infection at the constant rate of $\frac{s^i \alpha^i}{p^i}$ per dollar spent, the entire budget should be spent on the group for which this cost reduction per dollar spent is largest. As a result

$$C_{T-1}(I_{T-1}^1, \ldots, I_{T-1}^R) = \sum_{i=1}^R [s^i I_{T-1}^i + \delta s^i \Gamma^i(I_{T-1}^i)] - \delta \max(\frac{s^1 \alpha^1}{p^1}, \ldots, \frac{s^R \alpha^R}{p^R}).$$

Since the term in square brackets is a strictly increasing, strictly concave function of one variable (I^i_{T-1}), $C_{T-1}(I^1_{T-1}, \ldots, I^R_{T-1})$ is a strictly increasing, strictly concave function of R variables.

The proof that each of the minimized cost functions $\{C_t\}_{t=1}^{T-2}$ inherit these two properties proceeds exactly as before:

Theorem 5. *The cost function $C_t(I^1_t, \ldots, I^R_t)$ is strictly increasing for any $I^i_t \in [\frac{M_t}{p^i}, N^i]$ where $i = 1, \ldots, R$ and $t = 1, 2, ..., T-1$.*

Proof. We have already demonstrated the result for $t = T-1$. Now assume inductively that in any period $t < T-1$ the cost function starting in the subsequent period $C_{t+1}(I^1_{t+1}, \ldots, I^R_{t+1})$ is strictly increasing for any infection levels $(I^1_{t+1}, \ldots, I^R_{t+1})$ in the period $t+1$ rectangle. We will show that the cost function starting in period t given by $C_t(I^1_t, \ldots, I^R_t)$ is strictly increasing for any infection levels $(I^1_t, \ldots, I^R_t)$ in the period t rectangle.

Let

$$\Omega(I^1_t, \ldots I^R_t; F^1_t, \ldots, F^R_t) \equiv \sum_{i=1}^{R} s^i I^i_t + \delta C_{t+1}\left(\Gamma^1(I^1_t) - \alpha^1 F^1_t, \ldots, \Gamma^R(I^R_t) - \alpha^R F^R_t\right) \quad (14)$$

denote the objective function that the health authority minimizes. Note that by hypothesis $C_{t+1}(I^1_{t+1}, \ldots, I^R_{t+1})$ is strictly increasing for $(I^1_{t+1}, \ldots, I^R_{t+1})$ in the period $t+1$ rectangle and that $\Gamma^i(I^i_t)$ is strictly increasing in I^i_t for $i = 1, \ldots, R$. Thus, $\Omega(I^1_t, \ldots, I^R_t; F^1_t, \ldots, F^R_t)$ is strictly increasing for any infection levels $(I^1_t, \ldots, I^R_t)$ in the period t rectangle. Let $(\hat{F}^1_t, \ldots, \hat{F}^R_t)$ be the cost-minimizing treatment levels given initial infection levels $(\hat{I}^1_t, \ldots, \hat{I}^R_t)$ in the period t rectangle. Consider any smaller infection levels $(I^1_t, \ldots, I^R_t)$ in the period t rectangle such that $\hat{I}^i_t > I^i_t$ for some i and $\hat{I}^j_t = I^j_t$ for $j \neq i$, so that the infection level is strictly smaller in one component. Then the minimized discounted social cost from period t onward of entering period t with these strictly smaller infection levels is strictly smaller. That is,

$$\begin{aligned} C_t(\hat{I}^1_t, \ldots, \hat{I}^R_t) &= \Omega(\hat{I}^1_t, \ldots, \hat{I}^R_t; \hat{F}^1_t, \ldots, \hat{F}^R_t) \\ &> \Omega(I^1_t, \ldots, I^R_t, \hat{F}^1_t, \ldots, \hat{F}^R_t) \\ &\geq C_t(I^1_t, \ldots, I^R_t), \end{aligned}$$

where equality in the first line follows from the definition of the cost function, the inequality in the second line follows from the fact that the minimand is an increasing function of initial infection levels, and the inequality in the third line follows from cost minimization. So the cost function is strictly increasing for any $I^i_t \in [\frac{M_t}{p^i}, N^i]$ where i is any one of the R groups. □

Next we show that the cost function in every period is strictly concave over the rectangular region identified in Assumption 1.

Theorem 6. *The cost function $C_t(I^1_t, \ldots, I^R_t)$ is strictly concave for any $I^i_t \in [\frac{M_t}{p^i}, N^i]$ where $i = 1, \ldots, R$, and $t = 1, \ldots, T-1$.*

Proof. We have already established the result for $t = T-1$. Now assume inductively that in any period $t < T-1$ the cost function starting in the subsequent period $C_{t+1}(I^1_{t+1}, \ldots, I^R_{t+1})$ is strictly concave for any $(I^1_{t+1}, \ldots, I^R_{t+1})$ in the period $t+1$ rectangle. We will show that the cost function starting in period t given by $C_t(I_t^1, \ldots, I_t^R)$ is strictly concave for any $(I_t^1, \ldots, I_t^R)$ in the period t rectangle.

As before, let

$$\Omega(I_t^1, \ldots, I_t^R; F_t^1, \ldots, F_t^R) \equiv \sum_{i=1}^{R} s^i I_t^i + \delta C_{t+1}\left(\Gamma^1(I_t^1) - \alpha^1 F_t^1, \ldots, \Gamma^R(I_t^R) - \alpha^R F_t^R\right)$$

be the function that the health authority minimizes. Note that if $C_{t+1}(I_{t+1}^1, \ldots, I_{t+1}^R)$ is strictly concave in $(I_{t+1}^1, \ldots, I_{t+1}^R)$ for any infection levels in the period $t+1$ rectangle, then $\Omega(I_t^1, \ldots, I_t^R; F_t^1, \ldots, F_t^R)$ is strictly concave for any $(I_t^1, \ldots, I_t^R)$ in the period t rectangle, since $\Gamma^i(I_t^i)$ is strictly concave in I_t^i for $i = 1, \ldots, R$. By the definition of strict concavity,

$$\begin{aligned}\Omega(\lambda I_t^1 + (1-\lambda)\hat{I}_t^1, \ldots, \lambda I_t^R + (1-\lambda)\hat{I}_t^R; F_t^1, \ldots, F_t^R)& \\ > \lambda\Omega(I_t^1, \ldots, I_t^R; F_t^1, \ldots, F_t^R) + (1-\lambda)\Omega(\hat{I}_t^1, \ldots, \hat{I}_t^R; F_t^1, \ldots, F_t^R)&\end{aligned}$$

for any two distinct feasible starting infection levels $(I_t^1, \ldots, I_t^R)$ and $(\hat{I}_t^{\,1}, \ldots, \hat{I}_t^{\,R})$ in the period t rectangle, any constant $\lambda \in (0, 1)$, and any affordable treatment levels $(F_t^1, \ldots, F_t^R)$.

Now let

$$(F_t^{1\lambda}, \ldots, F_t^{R\lambda}) = \underset{F_t^i \geq 0 \text{ and } \sum_{i=1}^R p^i F_t^i \leq M_t}{\arg\min} \Omega(\lambda I_t^1 + (1-\lambda)\hat{I}_t^1, \ldots, \lambda I_t^R + (1-\lambda)\hat{I}_t^R; F_t^1, \ldots, F_t^R)$$

be the cost-minimizing treatment starting from initial infection levels $(\lambda I_t^1 + (1-\lambda)\hat{I}_t^1, \ldots, \lambda I_t^R + (1-\lambda)\hat{I}_t^R)$. Then

$$\begin{aligned}C_t(\lambda I_t^1 + (1-\lambda)\hat{I}_t^1, \ldots, \lambda I_t^R + (1-\lambda)\hat{I}_t^R)& \\ = \Omega(\lambda I_t^1 + (1-\lambda)\hat{I}_t^1, \ldots, \lambda I_t^R + (1-\lambda)\hat{I}_t^R; F_t^{1\lambda}, \ldots, F_t^{R\lambda})& \\ > \lambda\Omega(I_t^1, \ldots, I_t^R; F_t^{1\lambda}, \ldots, F_t^{R\lambda}) + (1-\lambda)\Omega(\hat{I}_t^1, \ldots, \hat{I}_t^R; F_t^{1\lambda}, \ldots, F_t^{R\lambda})& \\ \geq \lambda C_t(I_t^1, \ldots, I_t^R) + (1-\lambda)C_t(\hat{I}_t^1, \ldots, \hat{I}_t^R),&\end{aligned}$$

where the equality in the first line follows from the definition of the cost function, the inequality in the second line follows from strict concavity, and the inequality in the third line follows from cost minimization. Therefore, $C_t(I_t^1, \ldots, I_t^R)$ is strictly concave in $(I_t^1, \ldots, I_t^R)$. □

We now pull together the several results above to prove the key qualitative result of our analysis that in every period the health authority will focus its entire budget on one group or the other.

Theorem 7. *It is optimal in each period t $(t = 1, \ldots, T-1)$ to spend the entire budget on one group.*

Proof. Suppose the contrary—that in some period t it is optimal to *divide* the entire budget among two or more groups. Label any two of these favored groups as group A and group B. Consider two alternatives to this allegedly optimal treatment policy: (1) spending a little more on A (financing it entirely by reduced spending on B) or (2) spending a little less on A (and spending the money saved on group B). Since each of these two alternative treatment policies exhausts the budget, the hypothesized optimal policy can be regarded as a weighted average of the two alternatives for some weight $\lambda \in (0,1)$. Since $C_{t+1}(I^1_{t+1}, \ldots, I^R_{t+1})$ is strictly concave in the infection vector, $C_{t+1}(\Gamma(I^1_t) - \alpha^1 F^1_t, \ldots, \Gamma(I^R_t) - \alpha^R F^R_t)$ is strictly concave in the treatment policy, $(F^1_t, \ldots, F^R_t)$. It follows that C_{t+1} is strictly *quasiconcave* in the treatment policy. But this implies that the weighted average of these two alternative treatment policies results in a strictly higher cost than whichever of them achieves the smaller discounted social cost. But since that alternative treatment policy can be implemented using the period t budget, the weighted average is *a fortiori* suboptimal. Since this argument can be used to show that any treatment policy *dividing* the budget in period t ($t = 1, \ldots, T-1$) among two or more groups is suboptimal, it is optimal to *focus* the budget in period t on a single group. □

4 Extension to problem of finite-length with a wealth constraint rather than annual budgets

The above analysis shows that when resources cannot be transferred from one period to another, the health authority should spend its annual budget on one group. Here, we wish to establish that this result is not an artifact of our assumption of annual budgets but also may arise if a wealth constraint replaces the annual budget constraint. Assume the health authority is constrained by:

$$\sum_{t=1}^{T} \delta^{t-1} \left[p^A F^A_t + p^B F^B_t \right] \leq \bar{W}_1,$$

where $\bar{W}_1$ is the given initial wealth.

To facilitate the analysis we assume that in no period is wealth sufficient to treat every infected individual in either group. Mathematically, this assumption can be stated as follows.

Assumption 2. $N^i \geq I^i_t \geq \frac{\bar{W}_1 \delta^{1-t}}{p^i}$ *for* $i = A, B$ *for all* $t = 1, 2, ..., T$.

We will show that it is optimal for the health authority to allocate its entire budget to a single group. Again, the proof (although not the result) depends on concavity of the cost function. The analysis proceeds in two steps. First, we show that with just a single group the minimized cost function in every period is weakly concave in the overall wealth allocation. The proof depends critically on the above assumption. Given this concavity, it is then straightforward to show that with two groups the health authority will allocate all wealth to a single group.

Again, we work backwards from the end of the horizon. Spending anything in period T is foolish since the consequences are not felt until after the last period (T). Hence, $F_T = 0$ and the minimized cost in period T is $C_T(I_T, W_T) = sI_T$.[13] This in turn implies that, in

[13]Note that we have suppressed group subscripts, since in the first step we are analyzing the optimal allocation of wealth over time in a single group.

period $T-1$, saving has no value and any wealth remaining should be spent immediately. Thus, $F_{T-1} = \frac{W_{T-1}}{p}$ and the minimized cost at $T-1$ is $C_{T-1}(I_{T-1}, W_{T-1}) = sI_{T-1} + \delta\left(\Gamma(I_{T-1}) - \alpha\frac{W_{T-1}}{p}\right)$. This cost function is (1) strictly increasing in the inherited number of infections, (2) strictly decreasing in the inherited amount of wealth remaining, and (3) weakly concave in these two variables.[14] It is straightforward to verify that every prior minimized cost function ($t < T-1$) inherits these three properties. Define the minimized cost function in period t as follows:

$$C_t(I_t, W_t) = \min_{F_t \in [0, W_t/p]} sI_t + \delta C_{t+1}(\Gamma(I_t) - \alpha F_t, \delta^{-1}(W_t - F_t))$$

and consider the first property.

Theorem 8. *Under assumption 2 the cost function $C_t(I_t, W_t)$ is strictly increasing in I_t for all $t = 1, 2, ..., T-2$.*

Proof. As we showed above, $C_{T-1}(I_{T-1}, W_{T-1})$ is strictly increasing in I_{T-1} for all I_{T-1} in the relevant domain. Now assume inductively that in any period $t < T-1$ the next period's cost function $C_{t+1}(I_{t+1}, W_{t+1})$ is strictly increasing in I_{t+1} over the relevant domain. We will show that the cost function in period t inherits this property.

Let

$$\Omega(I_t, W_t, F_t) \equiv sI_t + \delta C_{t+1}(\Gamma(I_t) - \alpha F_t, \delta^{-1}(W_t - F_t)) \qquad (15)$$

be the function that the health authority minimizes in period t. Note that $C_{t+1}(I_{t+1}, W_{t+1})$ is strictly increasing in I_{t+1} by our inductive assumption and that $\Gamma(I_t)$ is strictly increasing in I_t. Therefore $\Omega(I_t, W_t, F_t)$ is strictly increasing in I_t by inspection. Let $\hat{F}_t$ be the treatment level that minimizes discounted social cost given inherited infection and wealth levels $(\hat{I}_t, \hat{W}_t)$, where $\hat{I}_t > I_t$ and $\hat{W}_t = W_t$. Then we have

$$\begin{aligned} C_t(\hat{I}_t, \hat{W}_t) &= \Omega(\hat{I}_t, \hat{W}_t, \hat{F}_t) \\ &> \Omega(I_t, \hat{W}_t, \hat{F}_t) \\ &\geq C_t(I_t, \hat{W}_t), \end{aligned}$$

where the equality in the first line follows from the definition of the cost function, the inequality in the second line follows because the minimand is an increasing function of the initial infection level, and the inequality in the third line follows from cost minimization. So the cost function in period t is strictly increasing in I_t. □

[14] Assumption 2 ensures that the health authority is never able to treat every infected individual, limiting the domain of the cost function (in inherited wealth and infection) to a region over which we can establish concavity. Without this assumption, the cost function is not concave in wealth. Consider period $T-1$. For $W_{T-1} \leq pI_{T-1}$, the first derivative of the cost function with respect to wealth is $-s\alpha\delta/p < 0$, since the extra wealth goes toward curing more people. For $W_{T-1} > pI_{T-1}$, however, the first derivative is zero, since every infected individual is already receiving treatment. There is therefore a kinked increase in the slope of the cost function at $W_{T-1} = pI_{T-1}$. This nonconcavity is passed to earlier periods and is inescapable. This nonconcavity would persist even if the health authority valued surplus wealth according to its dollar value, rather than zero. (This of course assumes that $s\alpha\delta > p$, which seems reasonable, else the health authority would do better to spend its budget on something other than treating disease.) Assumption 2 limits the relevant domain of the cost function so that this nonconcavity never presents.

A similar argument can be used to establish the second property that $C_t(I_t, W_t)$ inherits:

Theorem 9. *Under assumption 2 the cost function $C_t(I_t, W_t)$ is strictly decreasing in W_t for all $t = 1, 2, ..., T-1$.*

Proof. As we showed above, $C_{T-1}(I_{T-1}, W_{T-1})$ is strictly decreasing in W_{T-1} over the relevant domain, and we can again assume inductively that $C_{t+1}(I_{t+1}, W_{t+1})$ is strictly decreasing in W_{t+1} over the relevant domain. Because $C_{t+1}(I_{t+1}, W_{t+1})$ is strictly decreasing in W_{t+1}, by our inductive assumption, and because $W_{t+1} = \delta^{-1}(W_t - F_t)$ is strictly increasing in W_t, we know that $\Omega(I_t, W_t, F_t)$ in equation (15) is strictly decreasing in W_t by inspection. Let $\hat{W}_t < W_t$ and $\hat{I}_t = I_t$, and let $\hat{F}_t$ be optimal given these hatted state variables. Then the foregoing inequalities again hold for the same reasons as in Theorem 8, and they establish that the minimized cost function at t is strictly decreasing in W_t. □

The final property that $C_t(I_t, W_t)$ inherits is weak concavity.

Theorem 10. *Under assumption 2 the cost function $C_t(I_t, W_t)$ is weakly concave in I_t and W_t for all $t = 1, 2, ..., T-1$.*

Proof. As we showed above, $C_{T-1}(I_{T-1}, W_{T-1})$ is weakly concave in its two arguments.[15] Now assume inductively that $C_{t+1}(I_{t+1}, W_{t+1})$ is weakly concave. We will show that the cost function in period t inherits this property.

Recall the definition of $\Omega(I_t, W_t, F_t)$ in equation (15). Because $C_{t+1}(I_{t+1}, W_{t+1})$ is weakly concave, by induction, and because $I_{t+1} = \Gamma(I_t) - \alpha F_t$ is concave in I_t while $W_{t+1} = \delta^{-1}(W_t - F_t)$ is linear in W_t, we know the Ω function is weakly concave in (I_t, W_t).[16] Let F_t^λ be optimal given the inherited pair $\lambda(I_t, W_t) + (1-\lambda)(\hat{I}_t, \hat{W}_t)$.

It follows that

$$\begin{aligned} C_t(\lambda(I_t, W_t) + (1-\lambda)(\hat{I}_t, \hat{W}_t)) &= \Omega(\lambda(I_t, W_t) + (1-\lambda)(\hat{I}_t, \hat{W}_t), F_t^\lambda) \\ &\geq \lambda\Omega(I_t, W_t, F_t^\lambda) + (1-\lambda)\Omega(\hat{I}_t, \hat{W}_t, F^\lambda) \\ &\geq \lambda C_t(I_t, W_t) + (1-\lambda)C_t)(\hat{I}_t, \hat{W}_t), \end{aligned}$$

where the equality in the first line follows from the definition of the cost function, the inequality in the second line follows from weak concavity of the Ω function, and the inequality in the third line follows from cost minimization. □

[15] Again, assumption 2 is critical here. Without it, we do not even know that C_{T-1} is weakly concave, in which case the recursive proof that follows breaks down immediately.

[16] Observe that

$$\begin{aligned} C_{t+1}(\Gamma(\lambda I_t + (1-\lambda)\hat{I}_t) - \alpha F_t, \delta^{-1}(\lambda W_t + (1-\lambda)\hat{W}_t - F_t)) > \\ C_{t+1}(\lambda(\Gamma(I_t) - \alpha F_t) + (1-\lambda)(\Gamma(\hat{I}_t) - \alpha F_t), \lambda\delta^{-1}(W_t - F_t) + (1-\lambda)\delta^{-1}(\hat{W}_t - F_t) \geq \\ \lambda C_{t+1}(\Gamma(I_t) - \alpha F_t, \delta^{-1}(W_t - F_t)) + (1-\lambda)C_{t+1}(\Gamma(\hat{I}_t) - \alpha F_t, \delta^{-1}(\hat{W}_t - F_t)) \end{aligned}$$

for any two sets of distinct infection-wealth pairs (I_t, W_t) and $(\hat{I}_t, \hat{W})$ and any $\lambda \in (0, 1)$. The first inequality follows from $\Gamma^i(\cdot)$ strictly concave and $C_{t+1}(I_{t+1}, W_{t+1})$ strictly increasing in I_{t+1} (note the substitution $F_t = \lambda F_t + (1-\lambda)F_t$), while the second inequality follows from $C_{t+1}(I_{t+1}, W_{t+1})$ weakly concave in its arguments. Thus, the first and last lines together imply that $C_{t+1}(\Gamma(I_t) - \alpha F_t, \delta^{-1}(W_t - F_t))$ is weakly concave in (I_t, W_t). Thus, $\Omega(I_t, W_t, F_t)$ is also weakly concave in (I_t, W_t), since it is simply a linear combination of I_t and $C_{t+1}(\Gamma(I_t) - \alpha F_t, \delta^{-1}(W_t - F_t))$.

Given that every minimized cost function is weakly concave in the inherited wealth and the inherited number of infections, the health authority's problem in each period is weakly concave in F_t.[17]

Therefore in (F_t, F_{t+1}) space, the indifference curves are weakly concave and spending one's entire wealth in the current period or deferring the entire amount for the future (where it is optimal to spend it in one future period) always minimizes discounted social cost.

These conclusions hold both for group A and for group B.

We now pull together the results from this section to prove our key result that with multiple groups the health authority will allocate all wealth to a single group.

Theorem 11. *Under assumption 2, the health authority minimizing the discounted social cost of infection in two groups should exhaust all of its wealth treating a single group in a single period.*

Proof. Observe that minimizing infections is equivalent to allocating the budget optimally between the two groups in the first period and then allocating these group-specific budgets optimally over time for each group separately. Let W_1^i be the budget allocated to group i and let $C_1^i(I_1^i, W_1^i)$ (for $i = A, B$) be the minimized discounted cost from the initial period onward. Then the health authority's problem is to choose $W_1^A \geq 0$ and $W_1^B \geq 0$ to minimize

$$C_1^A(I_1^A, W_1^A) + C_1^B(I_1^B, W_1^B)$$

subject to the linear budget constraint $W_1^A + W_1^B \leq W_1$. We showed above that $C_1^i(I_1^i, W_1^i)$ is strictly decreasing in W_1^i for $i = A, B$. Hence, in (W_1^A, W_1^B) space, the preference direction is northeast, and the health authority should exhaust its overall budget. We also showed that $C_1^i(I_1^i, W_1^i)$ is weakly concave in W_1^i. Thus, $C_1^A(I_1^A, W_1^A) + C_1^B(I_1^B, W_1^B)$ is weakly concave in (W_1^A, W_1^B). Hence, in (W_1^A, W_1^B) space the indifference curves are weakly concave. That is, the set of wealth allocations (e.g., the origin) that have higher cost than the allocations on any given indifference curve is a weakly convex set. Thus, the cost-minimizing allocation is to focus the expenditure on one group. Having done this, the optimal allocation is to spend the entire budget in a single period, as we showed above.[18] □

An identical conclusion would follow for $R > 2$ groups.

[17] Observe that

$$\begin{aligned} C_{t+1}(\Gamma(I_t) - \alpha(\lambda F_t + (1-\lambda)\hat{F}_t), \delta^{-1}(W_t - (\lambda F_t + (1-\lambda)\hat{F}_t))) &= \\ C_{t+1}(\lambda(\Gamma(I_t) - \alpha F_t) + (1-\lambda)(\Gamma(I_t) - \alpha \hat{F}_t), \lambda\delta^{-1}(W_t - F_t) + (1-\lambda)\delta^{-1}(W_t - \hat{F}_t)) &\geq \\ \lambda C_{t+1}(\Gamma(I_t) - \alpha F_t, \delta^{-1}(W_t - F_t)) + (1-\lambda)C_{t+1}(\Gamma(I_t) - \alpha\hat{F}_t, \delta^{-1}(W_t - \hat{F}_t)) & \end{aligned}$$

for any two distinct treatment levels F_t and $\hat{F}_t$ and any $\lambda \in (0, 1)$. The inequality follows from $C_{t+1}(I_{t+1}, W_{t+1})$ weakly concave in its arguments. Thus, the first and last lines together imply that $C_{t+1}(\Gamma(I_t) - \alpha F_t, \delta^{-1}(W_t - F_t))$ is weakly concave in F_t. Thus, the minimization problem is weakly concave in F_t, since treatment has no effect on infections in period t.

[18] It actually can be shown that the cost function for allocating wealth optimally in one group over time is strictly concave in inherited wealth for all periods $t < T - 1$. This implies that, even in the case of two groups with identical parameters, it is optimal to allocate all wealth exclusively to one group or the other (i.e., the indifference curves are strictly concave). The proof of strict concavity requires showing that the cost function is strictly concave in period $T - 2$, which takes several pages. Thus, to save space, we only show the proof of weak concavity.

5 Conclusion

In this paper, we have considered the problem of a health authority attempting to minimize discounted social costs by allocating its budget among several distinct groups or regions to prevent the spread of a disease from sick to healthy people. Admittedly, our results depend on two strong assumptions: (1) that spending on treatment in a given region is not subject to increasing marginal costs over the relevant range and (2) that health budgets are tight. The first is a standard benchmark in the literature, even when ignoring health budgets. The second assumption seems to us plausible in many circumstances and makes the first assumption all the more plausible. Even given these assumptions, deriving conclusions using the standard optimal control or dynamic programming approaches is not feasible. However, our modified dynamic programming algorithm permits us to draw strong conclusions using familiar arguments for any finite number of groups.

The problem of optimal disease control in multiple groups or regions belongs to a more general class of problems that includes the spread of pests, crime (Philipson and Posner 1996), gang activity, and illegal drug use. In each of these cases, treating an afflicted individual not only reduces the costs he or she imposes on society but reduces imitation by others who may inflict additional social costs. Since different dynamics best describe the spread of these other social blights, a new model would be required. However, we anticipate that our methodology would continue to provide insights when applied in these other contexts.

References

Gersovitz, Mark and Jeffrey S. Hammer, "The Economical Control of Infectious Diseases," *The Economic Journal*, January 2004, *114* (492), 1–27.

_ **and** _ , "Tax/subsidy Policies Toward Vector-borne Infectious Diseases," *Journal of Public Economics*, April 2005, *89* (4), 647–674.

Goldman, Steven Marc and James Lightwood, "Cost Optimization in the SIS Model of Infectious Disease With Treatment," *Topics in Economic Analysis and Policy*, April 2002, *2* (1), Article 4.

Herrmann, Markus and Gérard Gaudet, "The Economic Dynamics of Antibiotic Efficacy Under Open Access," *Journal of Environmental Economics and Management*, May 2009, *57* (3), 334–350.

Lopez, Alan D., Colin D. Mathers, Majid Ezzati, Dean T. Jamison, and Christopher J.L. Murray, *Global Burden of Disease and Risk Factors*, New York: Oxford University Press, 2006.

Philipson, Thomas J. and Richard A. Posner, "The Economic Epidemiology of Crime," *Journal of Law and Economics*, October 1996, *39* (2), 405–433.

Revelle, Charles S., "The Economic Allocation of Tuberculosis Control Activities in Developing Nations." PhD dissertation, Cornell University 1967.

Rowthorn, Robert and Gardner M. Brown, "Using Antibiotics When Resistance is Renewable," in Ramanan Laxminarayan, ed., *Battling Resistance to Antibiotics and Pesticides: An Economic Approach*, Resources for the Future, 2003, pp. 42–62.

Rowthorn, Robert E, Ramanan Laxminarayan, and Christopher A Gilligan, "Optimal Control of Epidemics in Metapopulations," *Journal of the Royal Society Interface*, 2009, *6* (41), 1135–1144.

Sanders, J. L., "Quantitative Guidelines for Communicable Disease Control Programs," *Biometrics*, December 1971, *27*, 833–893.

Sethi, Suresh P., "Quantitative Guidelines for Communicable Disease Control Programs: A Complete Synthesis," *Biometrics*, December 1974, *30*, 681–691.

Smith, David L., Simon A. Levin, and Ramanan Laxminarayan, "Strategic Interactions in Multi-institutional Epidemics of Antibiotic Resistance," *Proceedings of the National Academy of Sciences*, February 2005, *102* (8), 3153–3158.

Appendix: A consequence of relaxing assumption 1

For $R = 2$ groups, assume without loss of generality that $\frac{\alpha^A s^A}{p^A} > \frac{\alpha^B s^B}{p^B}$. In that case, the budget at $T-1$ should be allocated to group A until every sick individual receives treatment. Suppose the number of sick individuals in group A is large enough, however, that treating them all is infeasible: $I^A_{T-1} \geq \frac{M_{T-1}}{p^A}$. It is then optimal to treat *none* of the sick in group B and, as derived in the text, the minimized cost function at $T-1$ is given by equation (8).

The right partial derivative at $I^A_{T-1} = \frac{M_{T-1}}{p^A}$ is therefore:

$$\frac{\partial C_{T-1}(I^A_{T-1}, I^B_{T-1})}{\partial I^A_{T-1}} = s^A + \delta \left[s^A \Gamma^{A\prime}(I^A_{T-1}) \right]. \tag{16}$$

We have proved that the cost function is strictly increasing and strictly concave for $I^A_{T-1} \geq \frac{M_{T-1}}{p^A}$.

If the number of sick individuals in group A is smaller, however, then all of the sick in group A can be treated with money left over. It is precisely this situation that we have ruled out with assumption 1. In this case, the health authority would begin to treat the sick of group B. For $I^A_{T-1} \leq \frac{M_{T-1}}{p^A}$ the minimized cost function is:

$$\begin{aligned} C_{T-1}(I^A_{T-1}, I^B_{T-1}) &= s^A I^A_{T-1} + s^B I^B_{T-1} \\ &+ \delta \left[s^A \Gamma^A(I^A_{T-1}) - \alpha^A s^A I^A_{T-1} + s^B \Gamma^B(I^B_{T-1}) - \alpha^B s^B \min\left(I^B_{T-1}, \left(\frac{M_{T-1} - p^A I^A_{T-1}}{p^B} \right) \right) \right]. \end{aligned}$$

The left-partial derivative at $I^A_{T-1} = \frac{M_{T-1}}{p^A}$ would therefore be

$$\frac{\partial C_{T-1}(I^A_{T-1}, I^B_{T-1})}{\partial I^A_{T-1}} = s^A + \delta \left[s^A \Gamma^{A\prime}(I^A_{T-1}) - \alpha^A s^A + \alpha^B s^B \frac{p^A}{p^B} \right]. \tag{17}$$

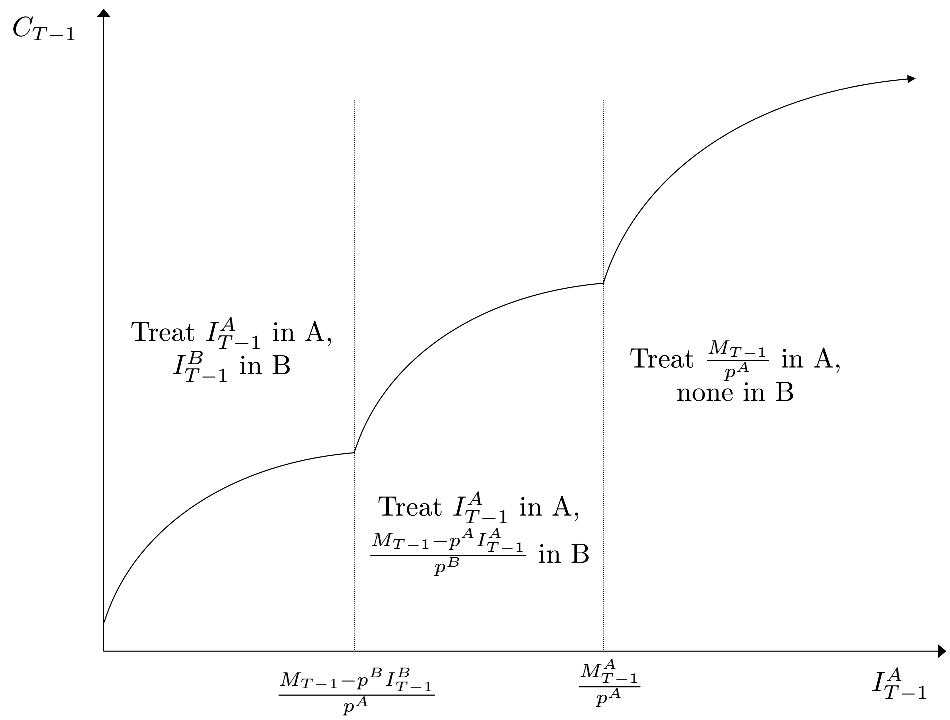

Figure 4: Kinks and curvature problems in the period $T-1$ cost function

Note: Figure shows a cross section of the period $T-1$ cost function, holding $I^B_{T-1} > 0$ constant, for all possible values of I^A_{T-1}. The cost function is not globally concave. See text for details.

Since the last two terms in square brackets are negative, the left partial derivative in (17) is strictly smaller than the right partial derivative in (16). Since Γ is strictly concave, the cost function is strictly concave for $I^A_{T-1} < \frac{M_{T-1}}{p^A}$. Using similar arguments, it can be shown that there is a second kink at $I^A_{T-1} = \frac{M_{T-1}-p^B I^B_{T-1}}{p^A}$, where the budget is just large enough to treat every sick person in both groups. Figure 4 illustrates these situations.

These kinks and curvature problems multiply as one proceeds backwards. For example, given initial conditions and optimal treatment in period $T-2$, it is possible that the resulting number of sick people in group A in period $T-1$ exactly matches the $T-1$ budget. This would generate a kink in the period $T-2$ cost function, even if the budget at $T-2$ was so tight as to make treating all the sick in group A infeasible. Of course, if the budget in period $T-2$ was large enough to treat all the sick in group A, then that would also generate a kink and convexity of the kind illustrated in Figure 4.

Such curvature problems "contaminate" each cost function generated in the backward recursion to earlier time periods. This contamination is unavoidable, so long as treating all the sick people in one or both groups remains a possibility in some future period. Using the standard dynamic programming algorithm, therefore, nothing could be said about the properties of the optimal policy rule, because the cost functions are not globally concave or convex. By making assumption 1 and altering the dynamic programming algorithm to take advantage of that assumption, we established that the cost function is always concave over the relevant range of infection levels. This permitted us to conclude that in each period the health authority should focus on the sick of a single group.