



ELSEVIER

Mathematical Biosciences 171 (2001) 33–58

**Mathematical  
Biosciences**

www.elsevier.com/locate/mbs

# Resource allocation for epidemic control over short time horizons

Gregory S. Zaric<sup>a,1</sup>, Margaret L. Brandeau<sup>b,\*</sup>

<sup>a</sup> *Ivey School of Business, University of Western Ontario, London, Ont., Canada N6A 3K7*

<sup>b</sup> *Department of Management Science and Engineering, Stanford University, Stanford, CA 94305, USA*

Received 17 March 1999; received in revised form 24 January 2001; accepted 8 February 2001

---

## Abstract

We present a model for allocation of epidemic control resources among a set of interventions. We assume that the epidemic is modeled by a general compartmental epidemic model, and that interventions change one or more of the parameters that describe the epidemic. Associated with each intervention is a ‘production function’ that relates the amount invested in the intervention to values of parameters in the epidemic model. The goal is to maximize quality-adjusted life years gained or the number of new infections averted over a fixed time horizon, subject to a budget constraint. Unlike previous models, our model allows for interacting populations and non-linear interacting production functions and does not require a long time horizon. We show that an analytical solution to the model may be difficult or impossible to derive, even for simple cases. Therefore, we derive a method of approximating the objective functions. We use the approximations to gain insight into the optimal resource allocation for three problem instances. We also develop heuristics for solving the general resource allocation problem. We present results of numerical studies using our approximations and heuristics. Finally, we discuss implications and applications of this work. © 2001 Elsevier Science Inc. All rights reserved.

*Keywords:* Epidemic control; Resource allocation; Optimization

---

## 1. Introduction

The control of the spread of infectious diseases such as influenza, malaria, tuberculosis, sexually transmitted diseases, and human immunodeficiency virus (HIV) is a challenging public health

---

\* Corresponding author. Tel.: +1-650 725 1623; fax: +1-650 725 8799.

*E-mail addresses:* gzaric@ivey.uwo.ca (G.S. Zaric), brandeau@stanford.edu (M.L. Brandeau).

<sup>1</sup> Tel.: +1-519 661 3415; fax: +1-519 661 3485.

problem. Characteristics related to the spread of an epidemic may differ across sub-populations within an at-risk population. Policy makers must decide how to allocate fixed budgets to the various sub-populations, through targeted interventions, to control the spread of the epidemic in the population as a whole.

Associated with each intervention one can think of a ‘production function’ that relates the amount invested in the intervention with some change in outcomes. Economists define the production function for a good as the amount of the good that can be produced from a set of inputs [1]. For epidemic control programs, outcomes may include infections averted, quality-adjusted life years (QALYs) gained, or other aggregate measures of community health. Some authors have defined the production functions for HIV prevention programs as the function that relates dollars invested in a program to the number of HIV infections averted [2–5]. The cost-effectiveness ratio, equal to the incremental cost of an intervention divided by the number of QALYs gained, is the standard unit of analysis in health care cost effectiveness [6]; such a ratio may represent a single point on a production function relating the amount invested to QALYs gained.

The effectiveness of an intervention is often measured by changes in parameters other than measures of community health such as QALYs gained or infections averted. For instance, studies of sexual risk counseling may measure effectiveness in terms of changes in condom use or number of new sexual partners per unit time [7–10]. Studies of medications may measure effectiveness in terms of changes in viral load or disease progression [11–15]. These changes may affect the course of an epidemic – and hence outcomes such as new infections or QALYs – but the relationship between the amount invested and these outcomes may be unknown. Thus, it is sometimes useful to view the production function as the relationship between the amount invested and the level of some intermediate variable that affects epidemic outcomes. Friedrich and Brandeau [16] and Richter et al. [17] demonstrated the application of such production functions in HIV resource allocation problems.

Different approaches to health care resource allocation problems have been proposed. Some authors have proposed linear or integer programming models [18–20] in which the objective is to maximize the sum of the benefits provided by several interventions subject to a budget constraint. These formulations assume that all combinations of interventions are feasible, that the benefits of any program are independent of the other programs selected, and that all programs are perfectly divisible with constant returns to scale (i.e., linear production functions) [21]. The optimal allocation under such assumptions is to select interventions in decreasing order of cost-effectiveness ratios until the entire budget has been consumed [20].

Similar mathematical programming approaches have been applied to resource allocation for the control of epidemics in general [22], and for HIV prevention in particular [3,4]. The goal is to allocate resources among non-interacting populations to minimize the number of new infections that occur over a finite time horizon, subject to a budget constraint. Richter et al. [22] assumed that each population is modeled by an SI epidemic model, and that a production function relates the amount invested in a population to the sufficient contact rate in that population. Thus, the relationship between the amount invested and the number of new infections is determined by the production function and the dynamics of the SI model in each population. Kaplan et al. [3,4] made no assumptions about the epidemic structure. They assumed that a production function relating the amount invested to infections averted can be constructed subjectively based on decision maker beliefs. Both approaches result in a knapsack problem that can be solved numerically.

Kahn [5] considered the impact of targeting HIV prevention resources to non-interacting populations. He considered an aggregate intervention with a fixed cost per person that reduces HIV incidence by a fixed percentage. He evaluated the impact of spending the entire budget on each single population. This approach implicitly assumes linear production functions, and considers only ‘all-or-nothing’ solutions.

Some resource allocation models aim to optimize some function of the equilibrium state of an epidemic [23,24]. However, not all epidemic models possess equilibrium states, and even if they do, the equilibrium state may not be reached for hundreds or thousands of years. Some researchers have simulated the trajectory of an epidemic to determine the best allocation of resources from among a limited set of alternatives [25–27]. This approach is limited by the small number of alternatives that can be considered. The best solution may be overlooked. Numerical techniques have also been used to determine the optimal allocation of resources [17,28,29], but it is not clear how well the numerical methods approximate the objective functions, nor if these techniques can be extended to more general models.

A related problem is to determine the optimal timing and intensity of interventions applied to a single population. The objective may be to minimize the cost of the interventions plus costs associated with infection subject to constraints on the level of disease prevalence at the end of the time horizon, or to minimize the cost of infection subject to budget constraints. Control theoretic techniques have been used to solve this type of problem. This approach has been used to study quarantine and removal programs [30,31], immunization programs [32,33], and other epidemic control programs [34–37].

In this paper we present a model for allocation of epidemic control resources that incorporates interacting populations and non-linear interacting production functions, and that does not require long time horizons. We formulate the model in Section 2. We show that an analytical solution to the model may be difficult or impossible to derive, even for simple cases. In Section 3 we derive a method of approximating the objective function. We use the approximations to gain insight into the optimal resource allocation for several special cases of the problem, and we develop heuristics for solving the general resource allocation problem. In Section 4 we present results of numerical studies using our approximations and heuristics. In Section 5 we discuss implications and applications of this work.

## 2. Resource allocation problem

We consider the problem of allocating resources among  $n$  interventions to optimize a function related to the spread of an epidemic over a time horizon of  $T$  years. All notation is summarized in Table 1. We assume that the epidemic is modeled by a compartmental epidemic model in which the population is divided into sub-populations (e.g., different risk groups) that are further divided by disease states (e.g., infected versus non-infected or differing stages of disease progression). Disease states are defined so that no disease state contains both infected and uninfected individuals. Each sub-population and disease-state pair defines a compartment; we assume that there are  $m$  such compartments. We assume that the epidemic is described by a system of non-linear differential equations, of the form  $x'_i(t) = f_i(x_1(t), x_2(t), \dots, x_m(t))$ ,  $i = 1, \dots, m$ . This is a common assumption (e.g., [38–40]) that allows us to model many important features of epidemics.

Table 1  
Summary of notation for RA

---

|                                  |                                                                                                           |
|----------------------------------|-----------------------------------------------------------------------------------------------------------|
| <i>Indices</i>                   |                                                                                                           |
| $g$                              | index for parameters, $g = 1, \dots, p$                                                                   |
| $h$                              | index for interventions, $h = 1, \dots, n$                                                                |
| $i, j$                           | indices for epidemic model compartments, $i, j = 1, \dots, m$                                             |
| $t$                              | index for time, $0 \leq t \leq T$                                                                         |
| $o$                              | index for objective functions, $o = q, n$                                                                 |
| <i>Decision variables</i>        |                                                                                                           |
| $v_h$                            | amount invested in intervention $h$ , $h = 1, \dots, n$                                                   |
| $\mathbf{v}$                     | $(v_1, v_2, \dots, v_n) =$ vector of investments                                                          |
| <i>Parameters</i>                |                                                                                                           |
| $P$                              | set of parameters specifying the compartmental epidemic model                                             |
| $p$                              | $ P $                                                                                                     |
| $n$                              | number of interventions                                                                                   |
| $m$                              | number of compartments in epidemic model                                                                  |
| $T$                              | length of time horizon                                                                                    |
| $w_g(\mathbf{v})$                | value of parameter $g$ , given investment vector $\mathbf{v}$ , $g = 1, \dots, p$                         |
| $\mathbf{w}(\mathbf{v})$         | $(w_1(\mathbf{v}), w_2(\mathbf{v}), \dots, w_p(\mathbf{v}))$                                              |
| $\mathbf{V}$                     | feasible region for $\mathbf{v}$                                                                          |
| $B$                              | total available budget                                                                                    |
| $r$                              | discount rate                                                                                             |
| $q_i$                            | quality adjustment for life years lived by individuals in compartment $i$ , $i = 1, \dots, m$             |
| $\lambda_{ij}$                   | rate of sufficient contact between individuals in compartments $i$ and $j$ , $i, j = 1, \dots, m$         |
| <i>Calculated quantities</i>     |                                                                                                           |
| $x_i(t, \mathbf{w}(\mathbf{v}))$ | size of compartment $i$ at time $t$ , given parameter values $\mathbf{w}(\mathbf{v})$ , $i = 1, \dots, m$ |
| $H_o(\mathbf{v})$                | value of objective function $o$ given investment vector $\mathbf{v}$                                      |

---

Interventions are assumed to change one or more of the parameters that describe the epidemic. For example, interventions may reduce the rate of sufficient contacts between infected and uninfected individuals, cause high-risk individuals to move to lower-risk populations, or change the rate of disease progression. Interventions may be targeted to specific population groups, or may reach the entire population. Interaction between interventions may also occur: for example, a general education campaign may increase awareness of a disease, and thus increase the effectiveness of other prevention programs.

Let  $P$  be a set of parameters that specifies the compartmental epidemic model, and let  $p = |P|$ , the number of parameters. Such parameters may include the rates of sufficient contact between members of different compartments, mortality rates, replacement rates, migration rates between risk groups, and rates of disease progression.

Let  $B$  be the decision maker's budget. Let  $v_h$ ,  $0 \leq v_h \leq V_h$  be the amount of money invested in intervention  $h$ ,  $h = 1, \dots, n$ , and let  $\mathbf{v} = (v_1, v_2, \dots, v_n)$  be the vector of investments. We assume that  $v_h \geq 0$ ; if we have a constraint  $v_h \geq c$ , we let  $v'_h = v_h - c$  and use  $v'_h$  as our decision variable with the constraint  $v'_h \geq 0$ . We let  $\mathbf{V}$  denote the set of feasible values for  $\mathbf{v}$ . We assume that the resource allocation decision is made at time zero and that interventions must be paid for at time zero.

Associated with each intervention is a production function that relates the amount invested in the intervention to the values of the parameters in  $P$ . Let  $w_g(\mathbf{v})$ ,  $g = 1, \dots, p$ , be the value of parameter  $g$  associated with the resource allocation  $\mathbf{v} = (v_1, v_2, \dots, v_n)$ , and let  $\mathbf{w}(\mathbf{0}) = (w_1(0), w_2(0), \dots, w_p(0))$  be the vector of initial parameter values. If a parameter is not affected by allocation decisions, then  $w_g(\mathbf{v}) = w_g(\mathbf{0})$ . We define  $\mathbf{w}(\mathbf{v}) : \mathcal{R}^n \rightarrow \mathcal{R}^p$  by  $\mathbf{w}(\mathbf{v}) = (w_1(\mathbf{v}), w_2(\mathbf{v}), \dots, w_p(\mathbf{v}))$ . We assume that interventions take effect instantaneously and that their effects last until the end of the time horizon.

Let  $x_i(t, \mathbf{w}(\mathbf{v}))$  be the size of compartment  $i$  at time  $t$ ,  $i = 1, \dots, m$ , given production function  $\mathbf{w}(\cdot)$ , and investments  $\mathbf{v}$ . We assume that the initial compartment sizes,  $x_i(0, \mathbf{w}(\mathbf{0}))$ , are known for all  $i$ . To simplify notation we write  $x_i(t)$  for  $x_i(t, \mathbf{w}(\mathbf{v}))$  unless we wish to emphasize the impact of the intervention on compartment sizes.

We let  $H_o(\mathbf{v})$  be the objective function for the resource allocation problem, where  $o$  is an index over possible objective functions. The resource allocation problem, RA, is

$$\text{RA : } \underset{\mathbf{v}}{\text{optimize}} H_o(\mathbf{v}) \tag{1}$$

$$\text{s.t. } \sum_{h=1}^n v_h \leq B, \tag{2}$$

$$\mathbf{v} \in \mathcal{V}. \tag{3}$$

We assume that  $H_o(\mathbf{v})$  is monotonically increasing in  $v_h$  in maximization problems, and monotonically decreasing in  $v_h$  in minimization problems; allocating more money to any intervention can only improve the objective function. Thus, the budget constraint is binding in the optimal solution and the total cost is  $B$  for all allocations.

We consider as objective functions total QALYs gained (which includes life years gained as a special case) and cumulative number of infections averted. QALYs gained is the recommended measure of benefit in applications of cost-effectiveness analysis to health and medicine [6]; we also consider cumulative infections averted, similar to some researchers in HIV prevention [4,22]. We discount benefits, at rate  $r$ , to reflect the economic principle that benefits that accrue sooner are preferable to benefits that accrue later. If the time horizon is short, it may be acceptable to ignore discounting.

The number of QALYs gained equals the number of QALYs experienced in the post-intervention health states minus the number of QALYs that would have been experienced in the absence of the intervention. Since the number of QALYs experienced in the absence of interventions is constant, we can ignore it and use total QALYs experienced in the post-intervention health states as the objective function. We do not attach a value to health states at the end of the time horizon, and consider only QALYs experienced until time  $T$ . This is equivalent to assuming a ‘salvage value’ of zero for all lives at the end of the time horizon, and ignores differences in QALYs that are accrued beyond time  $T$  as a result of current allocation decisions. Let  $q_i$  be the quality adjustment for life years lived by individuals in compartment  $i$ ,  $0 \leq q_i \leq 1$ . The discounted total number of QALYs experienced until time  $T$ , given investment  $\mathbf{v}$ , is

$$H_q(\mathbf{v}) = \text{QALY}(\mathbf{v}) = \sum_{i=1}^m q_i \int_{t=0}^T e^{-rt} x_i(t, \mathbf{w}(\mathbf{v})) dt. \tag{4}$$

An expression for discounted total life years experienced is obtained by setting  $q_i = 1$  for all  $i$ .

Maximizing the cumulative number of infections averted is equivalent to minimizing the cumulative number of new infections that occur. Let  $\lambda_{ij} \in P$  be the rate of sufficient contacts between members of compartments  $i$  and  $j$ . We assume that  $\lambda_{ij} = 0$  for all  $i, j$  such that  $i$  is in the same disease stage as  $j$  (i.e., the infection can only be passed from an infected to an uninfected individual). Assuming random mixing, the discounted number of new infections that occur up to time  $T$ , given investment  $\mathbf{v}$ , is

$$H_n(\mathbf{v}) = \text{INF}(\mathbf{v}) = \int_{t=0}^T e^{-rt} \left[ \frac{\sum_{i=1}^m x_i(t, \mathbf{w}(\mathbf{v})) \sum_{j=1}^m \lambda_{ij} x_j(t, \mathbf{w}(\mathbf{v}))}{\sum_{i=1}^m x_i(t, \mathbf{w}(\mathbf{v}))} \right] dt. \quad (5)$$

As for the QALYs objective function, we do not attach terminal values to health states at time  $T$ .

### 3. Analysis of the resource allocation problem

The epidemic model determines the compartment size functions,  $x_i(t)$ , and thus determines the form of the objective function and budget constraint. Closed-form expressions for  $x_i(t)$  are only known for the SI model and closely related variants [39]. Thus the solution of RA is, in general, very challenging.

We begin this section by specifying RA for the SI epidemic model. We show that, even for this simple case, RA is complex. We develop an approximation method to aid in solving RA. We illustrate our method for several special instances of RA and we characterize the optimal solutions to the approximated problems. We also present heuristics for solving more general instances of RA.

#### 3.1. A simple epidemic model

The SI epidemic model with replacement [39] is

$$x_1'(t) = \delta - \lambda x_1(t)x_2(t) - \delta x_1(t), \quad (6)$$

$$x_2'(t) = \lambda x_1(t)x_2(t) - \delta x_2(t), \quad (7)$$

$$x_1(t) + x_2(t) = 1. \quad (8)$$

The term  $x_1(t)$  represents the proportion of the population that is susceptible at time  $t$ ,  $x_2(t)$  represents the proportion that is infected at time  $t$ ,  $\delta$  is the replacement rate, and  $\lambda$  is the sufficient contact rate. The total population size is constant. If  $N$  is the population size, then  $Nx_1(t)$  and  $Nx_2(t)$  are the numbers of susceptibles and infecteds in the population at time  $t$ . The solution to this model is

$$x_2(t) = \begin{cases} \frac{x_2(0)(\lambda - \delta)e^{(\lambda - \delta)t}}{\lambda x_2(0)(e^{(\lambda - \delta)t} - 1) + (\lambda - \delta)}, & \lambda \neq \delta, \\ \frac{x_2(0)}{\lambda x_2(0)t + 1}, & \lambda = \delta, \end{cases} \quad (9)$$

where  $x_2(0)$  is the proportion infected at time zero.

Suppose that investment in prevention reduces the sufficient contact rate  $\lambda$  and that the production function is given by  $\lambda(v)$ . Assuming no discounting ( $r = 0$ ), the total number of QALYs experienced by the population up to time  $T$ , given investment  $v$ , is

$$\begin{aligned}
 H_q(v) &= N \int_{t=0}^T [q_1 x_1(t) + q_2 x_2(t)] dt = N \int_{t=0}^T [q_1 + (q_2 - q_1)x_2(t)] dt \\
 &= \begin{cases} N \left( q_1 T + (q_2 - q_1) \frac{1}{\lambda(v)} \ln \left[ \frac{(\lambda(v) - \delta) + x_2(0)\lambda(v)(e^{(\lambda(v)-\delta) \cdot T} - 1)}{(\lambda(v) - \delta)} \right] \right), & \lambda(v) \neq \delta, \\ N \left( q_1 T + (q_2 - q_1) \frac{1}{\lambda(v)} \ln [\lambda(v)x_2(0)T + 1] \right), & \lambda(v) = \delta \end{cases}
 \end{aligned} \tag{10}$$

and the cumulative number of new infections up to time  $T$ , given investment  $v$ , is [22]

$$\begin{aligned}
 H_n(v) &= N \int_{t=0}^T \lambda(v)x_1(t)x_2(t) dt = N \int_{t=0}^T [x_2'(t) + \delta x_2(t)] dt \\
 &= \begin{cases} N \left( x_2(T) - x_2(0) + \frac{\delta}{\lambda(v)} \ln \left[ \frac{(\lambda(v) - \delta) + x_2(0)\lambda(v)(e^{(\lambda(v)-\delta) \cdot T} - 1)}{(\lambda(v) - \delta)} \right] \right), & \lambda(v) \neq \delta, \\ N \left( \frac{x_2(0)}{\lambda(v)x_2(0)T + 1} - x_2(0) + \ln[\lambda(v)x_2(0)T + 1] \right), & \lambda(v) = \delta. \end{cases}
 \end{aligned} \tag{11}$$

Assuming independent sub-populations, each modeled by an SI model, the objective function for RA is a sum of terms like (10) or (11), summed over different sub-populations. These expressions involve transcendental functions, are non-linear in  $\lambda(v)$ , and were derived assuming no discounting, which may make them inappropriate for long time horizons. It is not known whether closed-form expressions for  $H_q(\cdot)$  and  $H_n(\cdot)$  exist when discounting is considered. This motivates the development of approximations for the objective functions.

### 3.2. Approximating terms in RA

Consider the general resource allocation problem, RA. We approximate each compartment size function with a Taylor series polynomial expanded around  $t = 0$ . This is possible because we assumed that the epidemic is modeled by a system of differential equations, of the form  $x_i'(t) = f_i(x_1(t), x_2(t), \dots, x_m(t))$ , and because we assumed that compartment sizes at time zero are known. Integrating such a polynomial between  $t = 0$  and  $t = T$  yields a function of model parameters, which we substitute into the objective function of RA.

We will illustrate our approximation method using a second-order approximation; other orders of approximation could also be used. A second-order approximation for  $x_i(t)$  is

$$y_i(t) \cong x_i(0) + x_i'(0) \cdot t + x_i''(0) \cdot \frac{t^2}{2}. \tag{12}$$

We introduce the following notation for discounted polynomials in  $t$

$$D(j) = \int_{t=0}^T e^{-rt} t^j dt. \quad (13)$$

Using (12), the QALYs objective function can be approximated as

$$\begin{aligned} \text{QALY}(\mathbf{v}) &= \sum_{i=1}^m q_i \int_{t=0}^T x_i(t) dt \cong \sum_{i=1}^m q_i \int_{t=0}^T y_i(t) dt \\ &= \sum_{i=1}^m q_i [x_i(0) \cdot D(0) + x'_i(0) \cdot D(1) + x''_i(0) \cdot D(2)]. \end{aligned} \quad (14)$$

To approximate the new-infections objective function, we let  $\text{NI}_i(t)$  denote the instantaneous rate of new infection among members of compartment  $i$ . This can be written as

$$\text{NI}_i(t) = \sum_{j=1}^m \lambda_{ij} x_j(t) \frac{x_j(t)}{N(t)}, \quad (15)$$

where  $N(t)$  is the total population size at time  $t$ . The first and second derivatives of  $\text{NI}_i(t)$  are

$$\text{NI}'_i(t) = \sum_{j=1}^m \lambda_{ij} \left\{ x'_i(t) \frac{x_j(t)}{N(t)} + x_i(t) \frac{x'_j(t)}{N(t)} + x_i(t) x_j(t) \frac{N'(t)}{[N(t)]^2} \right\}, \quad (16)$$

$$\begin{aligned} \text{NI}''_i(t) &= \sum_{j=1}^m \lambda_{ij} \left\{ x''_i(t) \frac{x_j(t)}{N(t)} + 2x'_i(t) \frac{x'_j(t)}{N(t)} + x_i(t) \frac{x''_j(t)}{N(t)} + 2 \left\{ x'_i(t) x_j(t) + x_i(t) x'_j(t) \right\} \frac{N'(t)}{[N(t)]^2} \right. \\ &\quad \left. + x_i(t) x_j(t) \frac{N''(t)[N(t)]^2 - 2N(t)[N'(t)]^2}{[N(t)]^4} \right\}. \end{aligned} \quad (17)$$

We can substitute the known values of  $x_i(0)$  into (16) and (17) to obtain expressions for  $\text{NI}'_i(0)$  and  $\text{NI}''_i(0)$ . We then combine these expressions to form an approximate expression for the new-infections objective function

$$\text{INF}(\mathbf{v}) \cong \sum_{i=1}^m \left[ \text{NI}_i(0) \cdot D(0) + \text{NI}'_i(0) \cdot D(1) + \text{NI}''_i(0) \cdot D(2) \right]. \quad (18)$$

In (14) and (18) we derived objective function expressions in terms of model parameters and known constants ( $x_i(0)$ 's). We substitute in these known constant values and replace the model parameters by the appropriate functions  $w_g(\mathbf{v})$  to obtain an expression that is a function of the  $v_h$ 's only, and we use this in the optimization.

We now illustrate the approximation method and develop results characterizing the optimal resource allocation (for the approximated problem) for several special cases of RA. All proofs are in Appendix A.

### 3.3. Analysis of a four-compartment model

Consider an epidemic model with two populations and two disease stages ( $m = 4$ ), as shown in Fig. 1. Individuals have sufficient contacts only with other members of their population. However,

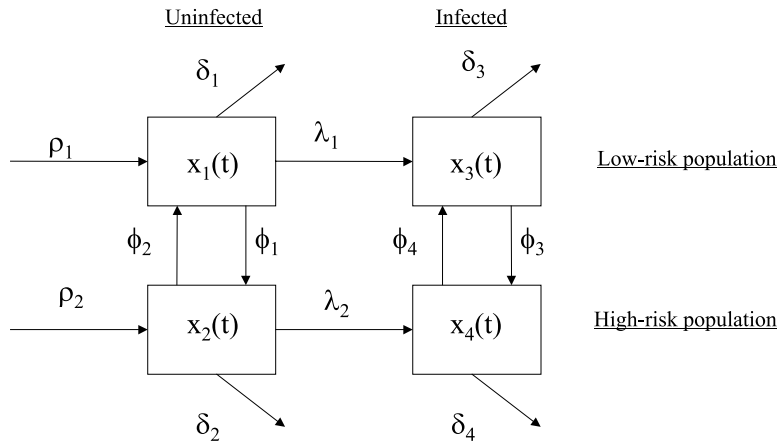


Fig. 1. A two-stage epidemic model with migration.

they may migrate between the two populations. Such a model may represent a population of injection drug users (IDUs) and a population of individuals who do not inject drugs (non-IDUs), where most risky contacts happen within the same population group. For example, injection sufficient contacts happen only among IDUs, and a significant proportion of IDU sexual contacts tend to be with other IDUs [41,42]. This model makes the strong simplifying assumption that no sufficient contacts occur across populations (e.g., between IDUs and non-IDUs). The epidemic model equations are

$$x'_1(t) = \rho_1(x_1(t) + x_3(t)) - \delta_1x_1(t) + \phi_2x_2(t) - \phi_1x_1(t) - \lambda_1x_1(t) \frac{x_3(t)}{x_1(t) + x_3(t)}, \quad (19)$$

$$x'_2(t) = \rho_2(x_2(t) + x_4(t)) - \delta_2x_2(t) - \phi_2x_2(t) + \phi_1x_1(t) - \lambda_2x_2(t) \frac{x_4(t)}{x_2(t) + x_4(t)}, \quad (20)$$

$$x'_3(t) = -\delta_3x_3(t) + \phi_4x_4(t) - \phi_3x_3(t) + \lambda_1x_1(t) \frac{x_3(t)}{x_1(t) + x_3(t)}, \quad (21)$$

$$x'_4(t) = -\delta_4x_4(t) - \phi_4x_4(t) + \phi_3x_3(t) + \lambda_2x_2(t) \frac{x_4(t)}{x_2(t) + x_4(t)}. \quad (22)$$

The terms  $\rho_i$  represent new entrants into the population,  $\delta_i$  are exit rates (equal to the sum of the death rate plus all rates of migration out of the population),  $\phi_i$  are migration rates to other compartments, and  $\lambda_i$  are sufficient contact rates. A closed-form solution for the compartment size functions is not known. Hethcote [43] discusses equilibrium properties of a similar model (with additional constraints on the  $\phi_i$ s) but does not provide closed-form solutions for either the equilibrium or the compartment size functions.

We assume that the first population (compartments 1 and 3) is a low-risk group, and the second population (compartments 2 and 4) is a high-risk group: thus, we assume  $\lambda_1 \leq \lambda_2$ . We assume that infection decreases quality of life, and that quality of life is higher in the low-risk group. Let QH be a

quality-of-life multiplier associated with life years lived in the high-risk group, and let  $QI$  be a quality-of-life multiplier associated with infection,  $0 \leq QH, QI, \leq 1$ . We assume that  $q_1 = 1$ ,  $q_2 = QH$ ,  $q_3 = QI$ , and  $q_4 = QH \times QI$ . Thus,  $q_1 \geq q_2$  and  $q_3 \geq q_4$ . This may be reasonable if the high-risk group experiences comorbidities that the low-risk group does not experience: for instance, IDUs have higher rates of tuberculosis, mental illness, and overdose than non-IDUs [44–49].

We investigate two resource allocation problems using this epidemic model. In problem P1, the decision maker must allocate resources between two interventions that reduce sufficient contact rates ( $\lambda_1$  and  $\lambda_2$ ). The interventions can be thought of as risk-reduction programs targeted to the low- and high-risk populations, respectively. We assume production functions of the following form:

$$\lambda_1(v_1) = \lambda_1^0 - b_1 v_1, \quad (23)$$

$$\lambda_2(v_2) = \lambda_2^0 - b_2 v_2. \quad (24)$$

We let P1-Q and P1-I denote P1 with the QALYs and new-infections objective functions, respectively.

In problem P2, the decision maker must allocate resources between an intervention that reduces sufficient contacts in the high-risk population ( $\lambda_2$ ) and an intervention that causes high-risk individuals to change their behavior and migrate to the low-risk population (thus increasing  $\phi_2$ ). We assume the production functions are given by (24) and by

$$\phi_2(v_1) = \phi_2^0 + b_1 v_1. \quad (25)$$

We also assume that the rate of cessation of risky behavior is independent of infection status; thus,  $\phi_1 = \phi_2$ . We let P2-Q and P2-I denote P2 with the QALYs and new-infections objective functions, respectively.

In the optimal solutions, the budget constraint will be binding, so we substitute  $v_2 = B - v_1$ , yielding a univariate problem in  $v_1$ . We approximate the objective functions as described in the previous section, but with an additional approximation that the denominators in (16) and (17) (total population size) are constant. This allows us to ignore some terms in the approximated new-infections objective function and derive tractable problems. We refer to the objective functions as  $QALY(v_1)$  and  $INF(v_1)$ ; it is understood that we are referring to the approximations of  $QALY(v_1)$  and  $INF(v_1)$ .

Propositions 1–4 characterize the solution to P1 for the QALYs objective with second-order approximations and for the new-infections objective with first-order approximations.

**Proposition 1.** *For P1-Q with second-order approximations:*

- (i) *If  $x_1(0) < x_3(0)$  and  $x_2(0) < x_4(0)$ , then  $QALY(v_1)$  is convex. The optimal solution  $\mathbf{v}^*$  is either  $(0, B)$  or  $(B, 0)$ .*
- (ii) *If  $x_1(0) > x_3(0)$  and  $x_2(0) > x_4(0)$ , then  $QALY(v_1)$  is concave. The optimal solution  $\mathbf{v}^*$  is either  $(0, B)$ ,  $(B, 0)$ , or an interior point whose value is stated in Eq. (A.1).*

**Proposition 2.** *For P1-I with first-order approximations:*

- (i) *If  $x_1(0) < x_3(0)$  and  $x_2(0) < x_4(0)$ , then  $INF(v_1)$  is concave. The optimal solution  $\mathbf{v}^*$  is either  $(0, B)$  or  $(B, 0)$ .*

(ii) If  $x_1(0) > x_3(0)$  and  $x_2(0) > x_4(0)$ , then  $INF(v_1)$  is convex. The optimal solution  $\mathbf{v}^*$  is either  $(0, B)$ ,  $(B, 0)$ , or an interior point whose value is stated in Eq. (A.2).

One implication of Propositions 1 and 2 is that, if a decision maker is allocating resources between two populations with prevalence over 50% (assuming that the epidemic and production functions can be realistically modeled by (19)–(24)), then it is optimal to invest everything in one population and nothing in the other, regardless of the values of all other model parameters, and regardless of whether the QALYs or new-infections objective function is used. The population that receives investment varies depending on problem parameters. If prevalence in the high-risk population is below 50% (and, by assumption, is lower in the low-risk population), then it may be optimal to share resources between the populations, for both objective functions; for these cases, Propositions 3 and 4 characterize how the optimal solution to P1 varies with the model parameters.

**Proposition 3.** For P1-Q with second-order approximations, if  $x_1(0) > x_3(0)$  and  $x_2(0) > x_4(0)$ , then:

- (i)  $v_1^*$  is non-decreasing in  $\delta_1, \rho_2, \lambda_1^0, \phi_1$ , and  $B$ ;
- (ii)  $v_1^*$  is non-increasing in  $\delta_2, \rho_1$ , and  $\lambda_2^0$ ;
- (iii) if  $QI > 1/2$ ,  $v_1^*$  is non-decreasing in  $\delta_3$  and non-increasing in  $\delta_4$ ;
- (iv) if  $QI < 1/2$ ,  $v_1^*$  is non-increasing in  $\delta_3$  and non-decreasing in  $\delta_4$ ;
- (v)  $v_1^*$  may be increasing or decreasing in  $QH, QI, \phi_2, \phi_3$ , and  $\phi_4$ .

Proposition 3 requires the QALYs objective with a second-order approximation. Then the amount invested in the low-risk group ( $v_1^*$ ) will not decrease as the death rate among uninfected members of the low-risk group ( $\delta_1$ ) increases, and will not increase as the death rate among uninfected members of the high-risk group ( $\delta_2$ ) increases. The optimal investment  $v_1^*$  will not decrease as the initial sufficient contact rate in the low-risk group ( $\lambda_1^0$ ) increases, and will not increase as the initial sufficient contact rate in the high-risk group ( $\lambda_2^0$ ) increases. The relationship between  $v_1^*$  and the rates of new entry ( $\rho_1$  and  $\rho_2$ ) suggests that a greater portion of the budget should be directed to the population with the lower rate of new entrants. For modest quality adjustments for infection ( $QI > 1/2$ ), the amount invested in each population increases as the death rate among infected individuals in that population increases, and for large quality adjustments ( $QI < 1/2$ ) the amount invested in a population decreases as the death rate among infected individuals of that population increases.

**Proposition 4.** For P1-I with first-order approximations, if  $x_1(0) > x_3(0)$  and  $x_2(0) > x_4(0)$ , then:

- (i)  $v_1^*$  is non-decreasing in  $\delta_2, \delta_4, \phi_2, \phi_4, \rho_1, \lambda_1^0$ , and  $B$ ;
- (ii)  $v_1^*$  is non-increasing in  $\delta_1, \delta_3, \phi_1, \phi_3, \rho_2$ , and  $\lambda_2^0$ .

Proposition 4 says that when the new-infections objective is used with a first-order approximation, the amount invested in the low-risk population ( $v_1^*$ ) will not decrease if death rates in the high-risk population ( $\delta_2$  and  $\delta_4$ ) increase; if the rate at which individuals move from the high-risk population to the low-risk population ( $\phi_2$  and  $\phi_4$ ) increases; if the replacement rate in the low-risk population ( $\rho_1$ ) increases; or if the initial sufficient contact rate in the low-risk population ( $\lambda_1^0$ ) increases.

We now establish that, for problem P2-Q with second-order approximations, if both populations have prevalence below 50%, and if a large quality adjustment is made for life years lived in the high-risk group ( $QH < 1/3$ ), then the optimal solution may involve sharing of resources between the two interventions.

**Proposition 5.** *For P2-Q with second-order approximations:*

- (i) if  $x_2(0) > x_4(0)$ , and  $x_1(0) > x_3(0)$ , and  $QH < 1/3$ , then  $QALY(v_1)$  is concave;
- (ii) if  $QALY(v_1)$  is concave, then the optimal solution  $(v_1^*, v_2^*)$  is either  $(B, 0)$ ,  $(0, B)$ , or an interior point whose value is stated in Eq. (A.3).

### 3.4. Special case: linear epidemic approximations, targeted interventions with linear production functions

Another special case worthy of attention is the case of first-order compartment size approximations and linear production functions. We illustrate using the SI model ((6)–(8)) with non-interacting linear production functions; however, the analysis can be extended to more general epidemic models and linear interacting production functions. In fact, if production functions are linear, the analysis extends to any model of the form  $x'_i(t) = f_i(x_1(t), x_2(t), \dots, x_m(t))$ , where  $f_i(\cdot)$  is linear in the model parameters for all  $i$ . We show that for both the QALYs and new-infections objective functions, RA reduces to a knapsack linear program (LP) for which a greedy solution is optimal.

We assume  $k$  independent populations, and for each population  $i$ , we let index  $i$  denote the uninfected compartment and  $i + k$  denote the infected compartment. We use the methods of the previous sections to derive our approximations, and drop terms that do not depend on  $\lambda_i$  to obtain

$$QALY(\mathbf{v}) \cong \sum_{i=1}^k N_i(q_{i+k} - q_i)\lambda_i x_i(0)x_{i+k}(0)D(1), \tag{26}$$

$$INF(\mathbf{v}) \cong \sum_{i=1}^k N_i\lambda_i x_i(0)x_{i+k}(0)[D(0) + \delta \cdot D(1)]. \tag{27}$$

We assume that no interaction between interventions occurs, and consider linear production functions of the form  $\lambda_i = w_i(v_i) = \lambda_i^0 - a_i v_i$ , where  $a_i \geq 0$  is a cost-effectiveness parameter associated with intervention  $i$ . Thus,  $\mathbf{w}(\mathbf{v}) = (\lambda_1^0 - a_1 v_1, \lambda_2^0 - a_2 v_2, \dots, \lambda_k^0 - a_k v_k)$ . We assume that interventions cannot increase sufficient contact rates, and that sufficient contact rates cannot be reduced below a lower limit  $l_i$ , where  $l_i \leq \lambda_i^0$ . Under these assumptions, RA is a linear program, as follows:

$$\text{RA-LP : } \text{optimize } H_o(\mathbf{w}(\mathbf{v})) \tag{1}$$

$$\text{s.t. } \sum_{i=1}^k v_i \leq B, \tag{2}$$

$$\lambda_i \leq \lambda_i^0, \quad i = 1, \dots, k, \tag{28}$$

$$\lambda_i = \lambda_i^0 - a_i v_i, \quad i = 1, \dots, k, \tag{29}$$

$$\lambda_i \geq l_i, \quad i = 1, \dots, k, \tag{30}$$

$$v_i \leq V_i, \quad i = 1, \dots, k, \quad (31)$$

$$v_i \geq 0, \quad i = 1, \dots, k. \quad (32)$$

We simplify the LP by using (29) to remove  $\lambda_i$  everywhere. Substituting (29) into (28) yields (32). Substituting (29) into (30) yields  $v_i \leq (\lambda_i^0 - l_i)/a_i$ . We define  $L_i$  by

$$L_i = \min \left\{ \frac{\lambda_i^0 - l_i}{a_i}, V_i \right\}. \quad (33)$$

We substitute (29) into the objective functions and drop all terms not containing  $v_i$  (which are constants, and therefore do not affect the optimization). By rearranging terms, the objectives can be written as  $\max \sum_{i=1}^k d_i v_i$  where  $d_i$  is appropriately defined. RA becomes

$$\text{RA-LP:} \quad \max \sum_{i=1}^k d_i v_i \quad (34)$$

$$\text{s.t.} \quad \sum_{i=1}^k v_i \leq B, \quad (2)$$

$$v_i \leq L_i, \quad (35)$$

$$v_i \geq 0. \quad (32)$$

RA-LP is a knapsack LP. The optimal solution is of the greedy type, and can be easily stated. Assume that the populations have been sorted in decreasing order of  $d_i$  (i.e.,  $d_1 \geq d_2 \geq \dots \geq d_k$ ). Define an index  $D$  such that  $L_1 + \dots + L_D \leq B$ , and  $L_1 + \dots + L_D + L_{D+1} \geq B$ . The optimal solution is to allocate  $v_i = L_i$  to populations  $i = 1, \dots, D$ , to allocate the remaining resources to population  $D + 1$ , and to allocate nothing to populations  $D + 2, \dots, k$ . The greedy solution is equivalent to investing in the interventions in decreasing order of their approximated cost-effectiveness ratios.

As long as the budget is too small to allocate  $L_i$  to each intervention, some populations will not receive any resources. This all-or-nothing type of allocation is optimal regardless of which of the two objective functions is used, since the approximated problems have the same form – maximizing a linear function subject to the same constraints. When the objective function coefficients have the same relative ordering, the optimal solutions will be identical, but this is not the case in general. The results of the LP analysis suggest that if the decision maker is considering health effects only over a very short time horizon, then it is always best to target resources to high-risk populations. We discuss the issue of the time horizon further when we present the results of our numerical analyses.

The initial sufficient contact rates,  $\lambda_i^0$ , do not influence the ordering of the populations because  $\lambda_i^0$  does not appear in  $d_i$ . This may seem surprising, because the sufficient contact rate can be thought of as a measure of infection risk in a population, and the all-or-nothing allocation result ignores the relative risk levels of the populations. However, the initial sufficient contact rate is like a sunk cost. When production functions are linear, the current level of risk is not important to the decision maker. All that is important is the potential benefit that can be derived from incremental investments, and this benefit is determined by factors other than  $\lambda_i^0$ . If the prevention program

production functions are non-linear, however, the optimal investment may be affected by the initial sufficient contact rates.

We now provide some intuition about the formulas for the  $d_i$  coefficients in the case of no discounting. First consider terms in the QALYs objective function, (26). New infections in population  $i$  occur at rate  $\lambda_i x_i(t) x_{i+k}(t) \approx \lambda_i x_i(0) x_{i+k}(0)$  for small  $t$ , so the total number of new infections until time  $T$  is approximately  $\lambda_i x_i(0) x_{i+k}(0) T$ . The total change in QALYs is related to the cumulative number of new infections, which is approximately  $\lambda_i x_i(0) x_{i+k}(0) T^2/2$ . For every newly infected individual, the population loses  $q_i$  QALYs and gains  $q_{i+k}$  QALYs. Now consider terms in the new-infections objective function, (27). Since new infections occur at a linear rate in  $t$  (with the first-order approximation), the change in the number of individuals infected after  $T$  years is  $\lambda_i x_i(0) x_{i+k}(0) T$ . However, these are not the only new infections that occur. The new infections expression must also account for new infections that replace infected individuals who have died; this is approximately  $\delta_i \lambda_i x_i(0) x_{i+k}(0) T^2/2$ .

### 3.5. Heuristics for solving RA

We have determined the optimal resource allocation for a few special cases of the problem with simple epidemic models, linear production functions, and first- or second-order approximations. Extension to more sophisticated epidemic models or more general production functions may not be possible in all cases. Thus, we propose four heuristics for solving RA.

H1. Exhaustive search over discretized feasible region with simulated epidemic curves: For each intervention, discretize the range of feasible investment values, then enumerate the points over all vectors  $(v_1, \dots, v_{n-1}, v_n = B - \sum v_h)$  for which  $v_n \geq 0$ . To evaluate the objective function for each allocation, simulate the epidemic by dividing the time horizon into equal time steps, and utilizing the discrete-time version of the epidemic model. Choose the best resulting allocation.

H2. Dynamic programming with discretized feasible region: For each intervention, discretize the range of feasible investment values. Solve the resulting problem with dynamic programming.

H3. Linearized LP knapsack solution: Approximate each production function with a linear function. Construct first-order compartment size approximations. Determine the optimal (greedy) solution to the resulting knapsack LP.

H4. Penalty function method with randomly generated initial solutions: Replace the objective function with an approximation (of any order) developed earlier. Set the budget constraint to an equality. Select a random initial feasible solution. Use a penalty function method with quadratic penalty functions to improve the solution. Repeat for as many random initial solutions as desired.

H1 can be thought of as a systematic extension to the ‘enumeration over a finite set of alternatives’ approach discussed in the literature review. H1 does not require functional forms for the compartment size functions; H1 can be applied with any epidemic model that can be simulated. Details of a simulation approach that can be used to evaluate the objective function can be found in Zaric et al. [50]. Results obtained by H1 can be made arbitrarily accurate by increasing the number of intervals considered in the feasible region; however, the execution time grows exponentially with the number of intervals. It may be difficult to scale H1 to problems with large numbers of interventions unless good rules for pruning the search space are developed.

H2 requires a separable objective function. A separable objective function arises in problems with non-interacting populations and non-interacting production functions. For the special case

of the SI model and no discounting of benefits, the exact objective function ((10) or (11)) can be used. For more general cases, the approximations developed earlier can be used, or simulation can be used. H2 is similar to the method other authors have used to solve example problems [3].

H3 solves the LP knapsack problem that results from first-order approximations and linear production functions. Application of H3 is much easier than the other heuristics, and can be done without knowledge of optimization theory. H3 is similar to allocation heuristics that have been used by some HIV Prevention Community Planning Groups [3,51].

H4 can be applied to a broad range of problems, including those with interacting production functions and populations, and can be easily scaled to solve large problems. H4 can also be used with a simulated objective function, similar to H1. Details of penalty search procedures can be found elsewhere [52,53]. Even in its approximated form, RA is in general neither convex nor concave, so H4 is not guaranteed to yield an optimal solution.

#### 4. Numerical results

We performed computational tests to evaluate the quality of our approximations. First we tested how well the first- and second-order compartment size approximations predict actual compartment sizes, QALYs experienced, and new infections in a single population for the case of an SI epidemic model. Our goal in developing the approximation method is to be able to solve RA using more sophisticated epidemic models; we used the SI model for some computational analysis since a closed-form expression for the compartment size functions is available. For each trial we calculated the absolute percent error as

$$\text{error} = \left| \frac{(\text{approximated value}) - (\text{actual value})}{\text{actual value}} \right| \times 100\%.$$

Table 2 shows that first- and second-order approximations for all three quantities are very accurate for short time horizons (one to five years). The first-order approximation for QALYs is accurate (less than 1% error) for time horizons up to 10 years. The first-order approximations are

Table 2

Average absolute percent difference between actual and approximated compartment size functions, QALYs gained, and new infections, with no discounting. Each result is based on 1000 trials<sup>a</sup>

| Quantity         | Type of approximation | Time horizon |         |          |
|------------------|-----------------------|--------------|---------|----------|
|                  |                       | $T = 1$      | $T = 5$ | $T = 10$ |
| Compartment size | First-order           | 0.1          | 2.5     | 8.9      |
|                  | Second-order          | 0.0          | 0.4     | 2.7      |
| QALYs            | First-order           | 0.0          | 0.1     | 0.9      |
|                  | Second-order          | 0.0          | 0.0     | 0.2      |
| New infections   | First-order           | 0.0          | 0.8     | 2.9      |
|                  | Second-order          | 0.0          | 0.2     | 1.4      |

<sup>a</sup>We assumed an SI model with replacement. For each trial we created a problem by drawing  $\lambda$ ,  $\delta$ ,  $q_2$ , and  $x_2(0)$  randomly from uniform distributions. We assumed  $\lambda \sim U(0.01, 0.2)$ ,  $\delta \sim U(0.01, 0.1)$ ,  $q_2 \sim U(0.1, 1)$ , and  $x_2(0) \sim U(0.02, 0.6)$ .

not accurate for compartment size and new infections over long time horizons, but the second-order approximations are.

We tested the quality of our approximations when used in RA. We considered a problem with four independent populations, each modeled by an SI model, and four independent interventions. We considered linear and exponential production functions, low and high potential behavior change, and time horizons of  $T = 1$  and  $T = 10$  years.

We first compared H2 with the actual objective functions versus H2 with the second-order approximations of the objective functions. Table 3 shows that the difference between the actual and approximated objective function value was less than 1% in all cases for  $T = 1$ , but was as high as 19% for  $T = 10$ . One measure of accuracy of the approximations is the number of times H2 found the same allocation (i.e., same values for  $v_1, v_2, v_3, v_4$ ) for the approximated problem as for the actual problem. For the QALYs objective function, H2 found the same solution in more than 68% of the cases for  $T = 1$ , but in as few as 8% of cases for  $T = 10$ . All-or-nothing solutions were

Table 3

Comparison of H2 using approximated objective functions versus H2 using the actual objective function for instances of a problem with four independent populations, each modeled by an SI epidemic model, and four interventions. Each result is based on 100 trials<sup>a</sup>

| Objective      | Production functions | Potential behavior change <sup>b</sup> | Time horizon | Average absolute difference in objective function values (%) | Proportion of trials in which the same solution was obtained (%) | Proportion of trials in which an all-or-nothing solution was optimal (%) |
|----------------|----------------------|----------------------------------------|--------------|--------------------------------------------------------------|------------------------------------------------------------------|--------------------------------------------------------------------------|
| QALYs          | Linear               | Low                                    | $T = 1$      | 0.0001                                                       | 100                                                              | 100                                                                      |
|                |                      |                                        | $T = 10$     | 0.1200                                                       | 83                                                               | 97                                                                       |
|                | Exponential          | High                                   | $T = 1$      | 0.0008                                                       | 96                                                               | 96                                                                       |
|                |                      |                                        | $T = 10$     | 0.0658                                                       | 68                                                               | 90                                                                       |
|                |                      | Low                                    | $T = 1$      | 0.0001                                                       | 76                                                               | 77                                                                       |
|                |                      |                                        | $T = 10$     | 0.1271                                                       | 73                                                               | 84                                                                       |
| New infections | Linear               | High                                   | $T = 1$      | 0.0023                                                       | 100                                                              | 100                                                                      |
|                |                      |                                        | $T = 10$     | 18.2671                                                      | 79                                                               | 96                                                                       |
|                | Exponential          | Low                                    | $T = 1$      | 0.0145                                                       | 96                                                               | 96                                                                       |
|                |                      |                                        | $T = 10$     | 11.1067                                                      | 77                                                               | 90                                                                       |
|                |                      | High                                   | $T = 1$      | 0.0260                                                       | 88                                                               | 75                                                                       |
|                |                      |                                        | $T = 10$     | 17.4287                                                      | 76                                                               | 85                                                                       |
| Low            | $T = 1$              | 0.0153                                 | 92           | 15                                                           |                                                                  |                                                                          |
|                | $T = 10$             | 10.9988                                | 12           | 7                                                            |                                                                  |                                                                          |

<sup>a</sup> Values for  $\lambda_i, \delta_i, x_i(0)$ , and  $N_i$  were drawn from uniform distributions. The limits of the uniform distributions were set to simulate four differing risk levels (on average). We used second-order approximations everywhere. We divided the \$1 000 000 budget range into 500 discrete intervals of \$2,000 each.

<sup>b</sup> Behavior change refers to the possible change in parameter values. For ‘low’ potential behavior change we assumed that if the entire budget were invested in population  $i$  it would reduce  $\lambda_i$  by 10% of its original value ( $w(B) = 0.9 \times w(0)$ ). For ‘high’ potential behavior change we assumed that the entire budget could reduce  $\lambda_i$  by 50% of its original value.

optimal in more than 90% of cases with linear production functions, for both short and long time horizons. Such solutions were less likely to be optimal for longer time horizons or non-linear production functions.

We also compared H2 to H3. Table 4 shows that, for the QALYs objective function, the LP approximation (H3) yielded solutions with objective function value less than 0.2% above those obtained by the dynamic programming method (H2) for every problem instance. For the new-infections objective function, the LP approximation yielded solutions with objective function value up to 10% above those obtained by H2.

We next compared H1 to H4 for a problem with six epidemic model compartments (two populations and three disease stages) as illustrated in Fig. 2, four interventions, and with the goal of maximizing QALYs. The populations represent high- and low-risk groups. We assumed that the two populations interact: individuals in compartments 1 and 2 may become infected via sufficient contacts with individuals in compartments 3, 4, 5, and 6, and individuals may migrate between populations. We considered four interventions that would, respectively, increase the rate

Table 4

Comparison of H2 using approximated objective functions versus H3 using approximated objective functions for instances of a problem with four independent populations, each modeled by an SI epidemic model, and four interventions. Each result is based on 100 trials<sup>a</sup>

| Objective      | Production functions | Potential behavior change <sup>b</sup> | Time horizon | Average absolute difference in objective function values (%) | Proportion of trials in which the same solution was obtained (%) |
|----------------|----------------------|----------------------------------------|--------------|--------------------------------------------------------------|------------------------------------------------------------------|
| QALYs          | Linear               | Low                                    | $T = 1$      | 0.00                                                         | 66                                                               |
|                |                      |                                        | $T = 10$     | 0.04                                                         | 49                                                               |
|                | Exponential          | High                                   | $T = 1$      | 0.02                                                         | 63                                                               |
|                |                      |                                        | $T = 10$     | 0.15                                                         | 49                                                               |
|                |                      | Low                                    | $T = 1$      | 0.00                                                         | 47                                                               |
|                |                      |                                        | $T = 10$     | 0.03                                                         | 42                                                               |
| High           | $T = 1$              | 0.02                                   | 7            |                                                              |                                                                  |
|                | $T = 10$             | 0.20                                   | 5            |                                                              |                                                                  |
| New infections | Linear               | Low                                    | $T = 1$      | 1.26                                                         | 51                                                               |
|                |                      |                                        | $T = 10$     | 1.43                                                         | 48                                                               |
|                |                      | High                                   | $T = 1$      | 8.22                                                         | 56                                                               |
|                |                      |                                        | $T = 10$     | 7.92                                                         | 52                                                               |
|                | Exponential          | Low                                    | $T = 1$      | 1.30                                                         | 41                                                               |
|                |                      |                                        | $T = 10$     | 1.58                                                         | 36                                                               |
|                |                      | High                                   | $T = 1$      | 7.24                                                         | 12                                                               |
|                |                      |                                        | $T = 10$     | 10.15                                                        | 9                                                                |

<sup>a</sup> Values for  $\lambda_i$ ,  $\delta_i$ ,  $x_i(0)$ , and  $N_i$  were drawn from uniform distributions. The limits of the uniform distributions were set to simulate four differing risk levels (on average). We used second-order approximations everywhere. We divided the \$1 000 000 budget range into 500 discrete intervals of \$2000 each.

<sup>b</sup> Behavior change refers to the possible change in parameter values. For ‘low’ potential behavior change we assumed that if the entire budget were invested in population  $i$  it would reduce  $\lambda_i$  by 10% of its original value ( $w(B) = 0.9 \times w(0)$ ). For ‘high’ potential behavior change we assumed that the entire budget could reduce  $\lambda_i$  by 50% of its original value.

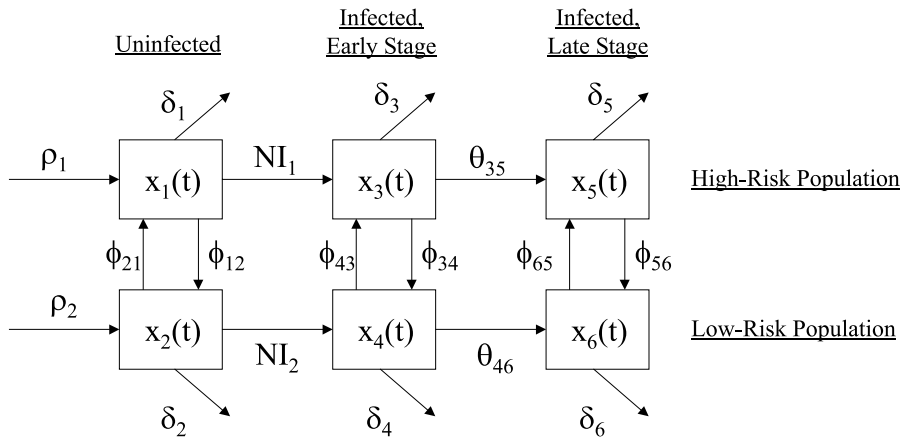


Fig. 2. A three-stage epidemic model with migration, disease progression, and cross infection.

of migration from the high- to low-risk group; decrease the sufficient contact rates of low-risk individuals; decrease the sufficient contact rates of high-risk individuals; and decrease the rate of disease progression of all individuals (from the infected early stage to the infected late stage). We defined  $w_g(\mathbf{v}) = M_g(\mathbf{v}) \times w_g(\mathbf{0})$ , where  $M_g(\mathbf{v})$  is an exponential function of the form

$$M_g(\mathbf{v}) = \alpha_g + \beta_g \exp \left( - \sum_{j=1}^n \gamma_{gj} v_j \right). \tag{36}$$

To fit the exponential production functions, we assumed that investing the entire budget in any intervention  $g$  would improve parameter  $g$  by  $x\%$  (i.e.,  $M_g = (1 - x/100)$  or  $M_g = (1 + x/100)$  depending on whether improvement corresponds to an increase or decrease in parameter values), and we assumed a maximum possible improvement of  $y\%$ . We modeled interaction between

Table 5

Comparison of H1 versus H4 for instances of a problem with six epidemic model compartments, four interventions, and the goal of maximizing QALYs. Each result is based on 100 trials<sup>a</sup>

| Problem | Time horizon | Average absolute difference H1 versus H4 (%) <sup>b</sup> |
|---------|--------------|-----------------------------------------------------------|
| 1       | $T = 1$      | 0.0000                                                    |
| 2       | $T = 1$      | 0.0025                                                    |
| 2       | $T = 5$      | 0.0072                                                    |
| 3       | $T = 5$      | 0.0001                                                    |

<sup>a</sup> Three separate data sets were used (Problems 1–3). All parameter values ( $\rho_i, \phi_{ij}, \theta_{ij}, \lambda_{ij}, \delta_i, \alpha_i, q_i$ , and  $x_i(0)$ ) were drawn from uniform distributions. The ranges were set to model a large low-risk population and a small high-risk population. In Problem 1,  $x$  and  $y$  were 0.25 and 0.5, and in Problems 2 and 3 they were 0.1 and 0.25. In Problems 1 and 2 we did not consider interacting production functions, but in Problem 3 we did. For H1 we divided the budget into 100 equal intervals of \$10 000 each. For each instance of H4 we considered 10 initial solutions.

<sup>b</sup> Calculated as the absolute percentage difference between the simulated objective function using the optimal solution obtained by H1 and the simulated objective function using the optimal solution obtained by H4. (Assumes that the simulated solution is a good approximation to the actual objective function.)

interventions by defining a parameter  $\chi_{gg'}$  that indicates the benefits obtained in intervention  $g$  from investment of \$1 in intervention  $g'$ . For example, if  $\chi_{12} = 0.1$ , then \$1 invested in intervention 2 also causes benefits equal to \$0.10 in intervention 1. We set  $\chi_{gg'} = 0.1$  for our experiment.

Table 5 shows that solutions obtained by H1 and H4 were very similar. In some cases H4 found better solutions than H1. We thus have some confidence in using H4 to solve larger problems. Solution times were long for H1 (on the order of 100–300 CPU s on a 400 MHz Intel Celeron processor). This makes it impractical to test H1 for longer time horizons, more interventions, or more budget intervals.

## 5. Discussion

Our model of resource allocation for epidemic control differs from existing models in that we explicitly allow for interacting populations and production functions, and we make no explicit assumptions about the shape of the production functions nor about the underlying epidemic structures. The general case of our formulation may yield a computationally intractable problem, even when a very simple form is assumed for the epidemic model. We developed approximation procedures that can yield a tractable problem, as well as solution heuristics. Our approximations yield good estimates of the objective function values, particularly when the time horizon is on the order of five years or less, and can be used as part of effective heuristics for solving large instances of RA.

In one special case, the problem reduces to a knapsack LP for which a greedy solution is optimal. Computational trials indicated that, when the QALYs objective function is used, the greedy solution may be appropriate in many cases; when the new-infections objective function is used, the greedy solution may be appropriate only when production functions are linear and time horizons are short (on the order of five years or less). In cases where the greedy heuristic is appropriate, the decision maker need not solve any optimization problems, but only needs to calculate the values of the objective function coefficients to determine the optimal allocation. For other situations, computational results indicate that heuristic H4 is likely to yield good solutions with reasonable computation time.

The methods we have presented can be adapted to alternative formulations of the problem. For example, guidelines for cost-effectiveness analyses in health and medicine call for the use of a societal perspective that measures all downstream costs and benefits [6]. This can be done by including in the model not only the immediate costs of the interventions, but also the future costs (or cost savings) associated with changes in health states. Prevention programs are likely to lead to future cost savings, whereas therapeutic interventions may sometimes lead to increases in future costs.

One way to include incremental these future costs (savings) in RA is to redefine the objective function as the net present total monetary benefit of the allocation decision: this is the net present total monetary benefit of QALYs gained minus the incremental costs (savings) minus the value of resources allocated. This is equivalent to minimizing the average cost-effectiveness ratio (total cost associated with investment  $v$  divided by total QALYs gained). The revised objective function can be written as

$$\begin{aligned} & \text{maximize } \mathbf{VQ} \times [\text{QALY}(\mathbf{v}) - \text{QALY}(\mathbf{0})] \\ & - \sum_{i=1}^m c_i \int_{t=0}^T e^{-rt} [x_i(t, \mathbf{w}(\mathbf{v})) - x_i(t, \mathbf{w}(\mathbf{0}))] - \sum_{h=1}^n v_h, \end{aligned} \quad (37)$$

where  $\mathbf{VQ}$  is the value that society places on QALYs gained and  $c_i$  is the economic cost per individual per year in compartment  $i$ . Although  $\mathbf{VQ}$  is difficult to specify, threshold values of \$20 000–\$50 000 are commonly accepted [54,55]. The modified resource allocation problem differs from RA only in the terms in the objective function. Both problems have a feasible region defined by a single budget constraint, so it is likely that the analysis and heuristics discussed for RA could be applied to the modified problem.

Our method for approximating new infections could provide a means of deriving production functions that directly link epidemic outcomes to investment (e.g., [3,4]). The production functions that we defined, in which changes in behavior are related to the amount invested, could serve as an intermediate step in calculating such production functions. This may be useful if decision makers can more readily think about changes in behavior than about epidemic outcomes.

We have assumed that the time horizon  $T$  is determined by external factors. For instance, if public health policy makers receive their budgets annually, then  $T$  might be one year in their planning model. However, for long term planning by some organizations, the choice of the appropriate time horizon may be important. Intuitively, one would expect that as the time horizon increases, the optimal mix of resources would shift away from interventions that increase QALYs and toward interventions that avert infections, since averted infections in the present lead to QALYs gained in the future. However, longer time horizons may lead to less reliable estimates of future benefits of interventions.

Our analysis has several limitations. We assumed that the budget is allocated at the start of the time horizon, and that intervention effects are instantaneous and constant. Zaric and Brandeau [56] considered an extension of RA in which resources are allocated over multiple discrete time periods, but assumed that intervention effects are instantaneous and constant. It is unlikely that an intervention could be implemented and be effective instantaneously. Moreover, some interventions may become less effective over time: for example, many studies have documented relapse to risky behavior after behavioral interventions aimed at epidemic control (e.g., [57–60]). Further work is needed to determine how these dynamic effects may affect the optimal allocation of resources. In addition, we assumed that the intervention production functions (e.g., the average levels of behavior change associated with a given investment in an intervention) are known and monotonic and do not depend on disease prevalence. Little data on such production functions exist; more work is needed to characterize these functions.

Our formulation of the resource allocation problem allows for interacting populations and interventions that affect more than one population. We have described approaches for directly solving this problem. An alternative approach is to approximate the problem to simplify the structure of the production functions or the epidemic model before solving. For example, it might be possible to model interacting epidemics among populations with independent epidemic models that use different parameters, or to model interacting production functions by an equivalent model with non-interacting production functions.

The resource allocation problem is complex. When solving, it is important to consider intervention production functions and non-linear epidemic growth. In some cases a greedy solution is

optimal. In other situations the approximations and heuristics developed in this paper can be used to determine the mix of resources that maximizes health benefits.

## Acknowledgements

This work was supported by the Societal Institute of the Mathematical Sciences (SIMS) through a grant from the National Institute on Drug Abuse (NIDA), National Institutes of Health (NIH) (Grant #DA R-01-09531-01A2).

## Appendix A.

### Value of interior point given in Proposition 1

When the optimal solution  $\mathbf{v}^*$  is an interior point, its value is given by

$$v_1 = \frac{b_1(x_2(0) + x_4(0))^2}{-2D(2)G_1} [A_1 + A_2 + A_3 + A_4 + A_5] + \frac{b_2(x_1(0) + x_3(0))^2}{-2D(2)G_1} [A_6 + A_7 + A_8 + A_9 + A_{10}],$$

$$v_2 = B - v_1,$$
(A.1)

where

$$A_1 = D1(q_1 - q_3)x_1(0)x_3(0)[x_1(0) + x_3(0)],$$

$$A_2 = D(2)x_1(0)x_3(0)[x_1(0) + x_3(0)] \\ \times \{-2q_1\delta_1 + q_3\delta_1 - q_1\delta_3 + 2q_3\delta_3 - 2q_1\phi_1 + q_2\phi_1 + q_3\phi_1 - q_1\phi_3 + 2q_3\phi_3 - q_4\phi_3\},$$

$$A_3 = 2D(2)x_1(0)x_3(0)[x_1(0) - x_3(0)](q_1 - q_3)\lambda_1^0,$$

$$A_4 = D(2)[x_1(0) + x_3(0)]^2x_3(0)\rho_1(q_1 - q_3),$$

$$A_5 = D(2)[x_1(0) + x_3(0)](q_1 - q_3)[x_1(0)x_4(0)\phi_4 + x_2(0)x_3(0)\phi_2],$$

$$A_6 = -D(1)(q_2 - q_4)x_2(0)x_4(0)[x_2(0) + x_4(0)],$$

$$A_7 = D(2)[x_2(0) + x_4(0)]x_2(0)x_4(0) \\ \times \{2q_2\delta_2 - q_4\delta_2 + q_2\delta_4 - 2q_4\delta_4 - q_1\phi_2 + 2q_2\phi_2 - q_4\phi_2 + q_2\phi_4 + q_3\phi_4 - 2q_4\phi_4\},$$

$$A_8 = 2D(2)x_2(0)x_4(0)[\lambda_2^0 - b_2B][x_2(0) - x_4(0)](q_4 - q_2),$$

$$A_9 = D(2)[x_2(0) + x_4(0)]^2x_4(0)\rho_2(q_4 - q_2),$$

$$A_{10} = D(2)[x_2(0) + x_4(0)](q_4 - q_2)[x_2(0)x_3(0)\phi_3 + x_1(0)x_4(0)\phi_1],$$

$$G_1 = b_1^2(q_1 - q_3)x_1(0)x_3(0)[x_2(0) + x_4(0)]^2[x_1(0) - x_3(0)] \\ + b_2^2(q_2 - q_4)x_2(0)x_4(0)[x_1(0) + x_3(0)]^2[x_2(0) - x_4(0)].$$

**Proof of Proposition 1.** (i) After algebraic manipulation, the second derivative of the approximated objective function can be written as

$$\frac{d^2}{dv_1^2} \text{QALY}(v_1) = \frac{-2D(2)}{(x_1(0) + x_3(0))^2(x_2(0) + x_4(0))^2} G_1.$$

Since  $q_1 \geq q_3$  and  $q_2 \geq q_4$ , we have  $(q_1 - q_3) > 0$  and  $(q_2 - q_4) > 0$ . If  $x_1(0) > x_3(0)$  and  $x_2(0) > x_4(0)$ , then  $G_1 \geq 0$ . If  $x_1(0) < x_3(0)$  and  $x_2(0) < x_4(0)$ , then  $G_1 \leq 0$ .

(ii) The maximum of a convex function over a convex region is an extreme point of the feasible region [53]. When maximizing a concave function, the optimal solution is either one of the two extreme points listed, or an interior point satisfying  $d\text{QALY}(v_1)/dv_1 = 0$ . After considerable algebraic manipulation, we see that such a point is given by (A.1).  $\square$

*Value of interior point given in Proposition 2.* When the optimal solution  $\mathbf{v}^*$  is an interior point, its value is given by

$$v_1 = \frac{b_1[x_2(0) + x_4(0)]^2[A_{11} + A_{12} + A_{13}] - b_2[x_1(0) + x_3(0)]^2[A_{14} + A_{15} + A_{16}]}{G_2}, \quad (\text{A.2})$$

$$v_2 = B - v_1,$$

where

$$A_{11} = D(0)[x_1(0) + x_3(0)]x_1(0)x_3(0),$$

$$A_{12} = D(1)[x_1(0) + x_3(0)] \times \{-x_1(0)x_3(0)[\delta_1 + \delta_3 - \rho_1 + \phi_1 + \phi_3] + x_3(0)^2\rho_1 + x_2(0)x_3(0)\phi_2 + x_1(0)x_4(0)\phi_4\},$$

$$A_{13} = 2D(1)x_1(0)x_3(0)\lambda_1^0[x_1(0) - x_3(0)],$$

$$A_{14} = D(0)[x_2(0) + x_4(0)]x_2(0)x_4(0),$$

$$A_{15} = D(1)[x_2(0) + x_4(0)]\{-x_2(0)x_4(0)[\delta_2 + \delta_4 - \rho_2 + \phi_2 + \phi_4] + x_4(0)^2\rho_2 + x_1(0)x_4(0)\phi_1 + x_2(0)x_3(0)\phi_3\},$$

$$A_{16} = 2D(1)x_2(0)x_4(0)[\lambda_2^0 - b_2B][x_2(0) - x_4(0)],$$

$$G_2 = 2D(1)\left\{b_1^2[x_2(0) + x_4(0)]^2x_1(0)x_3(0)[x_1(0) - x_3(0)] + b_2^2[x_1(0) + x_3(0)]^2x_2(0)x_4(0)[x_2(0) - x_4(0)]\right\}.$$

**Proof of Proposition 2.** (i) After algebraic manipulation, the second derivative of the approximated objective function can be written as

$$\frac{d^2}{dv_1^2} \text{INF}(v_1) = \frac{2D(1)}{(x_1(0) + x_3(0))^2(x_2(0) + x_4(0))^2} G_2.$$

The first term in the above equation is strictly positive. If  $x_1(0) > x_3(0)$  and  $x_2(0) > x_4(0)$ , then  $G_2$  is positive. If  $x_1(0) < x_3(0)$  and  $x_2(0) < x_4(0)$ , then  $G_2$  is negative.

(ii) When  $\text{INF}(v_1)$  is concave, the optimal solution,  $\mathbf{v}^*$ , is an extreme point of the feasible region: either  $(0, B)$ ,  $(B, 0)$ . When  $\text{INF}(v_1)$  is convex, the Karush–Kuhn–Tucker conditions may be satisfied either at  $(0, B)$ ,  $(B, 0)$ , or  $\mathbf{v}^* = (v_1, B - v_1)$  as given in (A.2).  $\square$

**Proof of Proposition 3.** Under the assumptions of this proposition, the denominator of (A.1) is strictly negative. The results follow from the definitions of  $A_1 - A_{10}$ .  $\square$

**Proof of Proposition 4.**  $v_1$  is given by (A.2). Since  $x_1(0) > x_3(0)$  and  $x_2(0) > x_4(0)$ , we have  $G_1 > 0$ . The results follow from the definitions of  $A_{11} - A_{16}$ .  $\square$

*Value of interior point given in Proposition 2.* When the optimal solution  $\mathbf{v}^*$  is an interior point, its value is given by

$$\begin{aligned} v_1 &= \frac{b_1(x_2(0) + x_4(0))}{G_3 + G_4 + G_5} [A_{17} + A_{18} + A_{19} + A_{20} + A_{21} + A_{22}] \\ &\quad - \frac{b_2(x_1(0) + x_3(0))}{G_3 + G_4 + G_5} [A_{23} + A_{24} + A_{25} + A_{26} + A_{27}], \\ v_2 &= B - v_1, \end{aligned} \tag{A.3}$$

where

$$\begin{aligned} A_{17} &= D(1)(x_1(0) + x_2(0))(x_2(0) + x_4(0))((q_1 - q_2)x_2(0) + (q_3 - q_4)x_4(0)), \\ A_{18} &= D(2)(x_1(0) + x_3(0))(x_2(0) + x_4(0)) \\ &\quad \times [-q_1x_1(0)\delta_1 + x_2(0)\delta_2(-q_1 + 2q_2) - q_3x_4(0)\delta_3 + x_4(0)\delta_4(-q_3 + 2q_4)], \\ A_{19} &= D(2)[x_1(0) + x_3(0)][x_2(0) + x_4(0)] \\ &\quad \times [q_1\rho_1[x_2(0) + x_4(0)] + \rho_2(q_1 - 2q_2)(x_2(0) + x_4(0))], \\ A_{20} &= D(2)x_2(0)x_4(0)[x_1(0) + x_3(0)](b_2B - \lambda_2^0)(q_1 - 3q_2 - q_3 + 3q_4), \\ A_{21} &= D(2)[x_2(0) + x_4(0)]\lambda_1(q_3 - q_1)[x_2(0)x_3(0) + x_1(0)x_4(0)], \\ A_{22} &= D(2)[x_1(0) + x_3(0)][x_2(0) + x_4(0)]\{\phi_1(q_1 - q_2)[x_1(0) - x_2(0)] \\ &\quad + 2\phi_2^0[x_2(0)(q_2 - q_1) - x_4(0)(q_4 - q_3)] + \phi_3(q_3 - q_4)[x_3(0) - x_4(0)]\}, \\ A_{23} &= D(1)[x_2(0) + x_4(0)]x_2(0)x_4(0)(q_4 - q_2), \\ A_{24} &= D(2)[x_2(0) + x_4(0)]x_2(0)x_4(0)\{\delta_2(2q_2 - q_4) - \delta_4(2q_4 - q_2)\}, \\ A_{25} &= D(2)[x_2(0) + x_4(0)]^2x_4(0)\rho_2(q_4 - q_2), \\ A_{26} &= 2D(2)x_2(0)x_4(0)[b_2B - \lambda_2^0](q_2 - q_4)[x_2(0) - x_4(0)], \end{aligned}$$

$$A_{27} = D(2)[x_2(0) + x_4(0)]\{\phi_1 x_1(0)x_4(0)(q_4 - q_2) + \phi_3 x_2(0)x_3(0)(q_4 - q_2) \\ \times \phi_2^0 x_2(0)x_4(0)(-q_1 + 3q_2 + q_3 - 3q_4)\},$$

$$G_3 = -2b_1^2 D(2)[x_1(0) + x_3(0)][x_2(0) + x_4(0)]^2\{(q_1 - q_2)x_2(0) + (q_3 - q_4)x_4(0)\},$$

$$G_4 = -2b_1 b_2 D(2)[x_2(0) + x_4(0)][x_1(0) + x_3(0)]x_2(0)x_4(0)\{q_1 - 3q_2 - q_3 + 3q_4\},$$

$$G_5 = -2b_2^2 D(2)[x_1(0) + x_3(0)]x_2(0)x_4(0)(q_2 - q_4)[x_2(0) - x_4(0)].$$

**Proof of Proposition 5.** (i) The second derivative of the approximated objective function is

$$\frac{d^2}{dv_1^2} \text{QALY}(v_1) = \frac{-2D(2)}{[x_2(0) + x_4(0)]^2} \{G_6 + G_7 + G_8\},$$

where

$$G_6 = b_2^2 (q_2 - q_4)x_2(0)x_4(0)[x_2(0) - x_4(0)] \geq 0,$$

$$G_7 = b_1^2 [x_2(0) + x_4(0)]^2 [(q_1 - q_2)x_2(0) + (q_3 - q_4)x_4(0)] \geq 0,$$

$$G_8 = b_1 b_2 x_2(0)x_4(0)[x_2(0) + x_4(0)]^2 (q_1 - 3q_2 - q_3 + 3q_4) \\ = b_1 b_2 x_2(0)x_4(0)[x_2(0) + x_4(0)][(1 - \text{QI})(1 - 3\text{QH})] \geq 0.$$

(ii) When maximizing a concave function, the optimal solution is either one of the two extreme points listed, or an interior point satisfying  $d\text{QALY}(v_1)/dv_1 = 0$ . After considerable algebraic manipulation, we see that such a point is given by (A.3).  $\square$

## References

- [1] A. Mas-Colell, M.D. Whinston, J.R. Green, in: *Microeconomic Theory*, Oxford University, New York, 1995.
- [2] E.H. Kaplan, Economic analysis of needle exchange, *AIDS* 9 (1995) 1113.
- [3] E.H. Kaplan, Economic evaluation and HIV prevention community planning: a policy analyst's perspective, in: D.R. Holtgrave (Ed.), *Handbook of Economic Evaluation of HIV Prevention Programs*, Plenum, New York, 1998, p. 177.
- [4] E.H. Kaplan, H. Pollack, Allocating HIV prevention resources, *Socio-Econ. Plann. Sci.* 32 (1998) 257.
- [5] J.G. Kahn, The cost-effectiveness of HIV prevention targeting: how much more bang for the buck? *Am. J. Public Health* 86 (1996) 1709.
- [6] M.R. Gold et al., in: *Cost-Effectiveness in Health and Medicine*, Oxford University, New York, 1996.
- [7] J.B. Jemmott, L.S. Jemmott, G.T. Fong, Abstinence and safer sex HIV risk-reduction interventions for African American adolescents: a randomized controlled trial, *J. Am. Med. Assoc.* 279 (1998) 1529.
- [8] J.A. Kelly et al., The effects of HIV/AIDS intervention groups for high-risk women in urban clinics, *Am. J. Public Health* 84 (1994) 1918.
- [9] R.J. MacGowan et al., Sex, drugs and HIV counseling and testing: a prospective study of behavior-change among methadone-maintenance clients in New England, *AIDS* 11 (1997) 229.
- [10] N.S. Wenger et al., Effect of HIV antibody testing and AIDS education on communication about HIV risk and sexual behavior. A randomized, controlled trial in college students, *Ann. Intern. Med.* 117 (1992) 905.

- [11] R.S. Hogg et al., Improved survival among HIV-infected patients after initiation of triple-drug antiretroviral regimens, *Canadian Med. Assoc. J.* 160 (1999) 659.
- [12] S.M. Hammer et al., A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus and CD4 cell counts of 200 per cubic millimeter or less, *N. Engl. J. Med.* 337 (1997) 725.
- [13] R.M. Gulick et al., Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy, *N. Engl. J. Med.* 337 (1997) 734.
- [14] A.C. Collier et al., Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine, *N. Engl. J. Med.* 334 (1996) 1011.
- [15] P. Pezzotti et al., Increasing survival time after AIDS in Italy: the role of new combination antiretroviral therapies. Tuscany AIDS Study Group, *AIDS* 13 (1999) 249.
- [16] C.M. Friedrich, M.L. Brandeau, Using simulation to find optimal funding levels for HIV prevention programs with different costs and effectiveness, in: M. Katzper, J.G. Anderson (Eds.), *Proceedings of the 1998 Medical Sciences Simulation Conference, Society for Computer Simulation International, San Diego, 1998*, p. 58.
- [17] A. Richter, M.L. Brandeau, D.K. Owens, An analysis of optimal resource allocation for HIV prevention among injection drug users and nonusers, *Med. Decis. Making* 19 (1999) 167.
- [18] M.M. Chen, J.W. Bush, Maximizing health system output with political and administrative constraints using mathematical programming, *Inquiry* 13 (1976) 215.
- [19] A.A. Stinnett, A.D. Paltiel, Mathematical programming for the efficient allocation of health care resources, *J. Health Economics* 15 (1996) 641.
- [20] M. Weinstein, R. Zeckhauser, Critical ratios and efficient allocation, *J. Public Econom.* 2 (1973) 147.
- [21] M.C. Weinstein, From cost-effectiveness ratios to resource allocation: where to draw the line? in: F.A. Sloan (Ed.), *Valuing Health Care: Costs, Benefits, Effectiveness of Pharmaceuticals and Other Medical Technologies*, Cambridge University, New York, 1995, p. 77.
- [22] A. Richter, M.L. Brandeau, G.S. Zaric, Optimal resource allocation for epidemic control among multiple independent populations, Technical Report, Department of Management Science and Engineering, Stanford University, Stanford, CA, 2000.
- [23] H.W. Hethcote, J.W. Van Ark, Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs, *Math. Biosci.* 84 (1987) 85.
- [24] R.M. May, R.M. Anderson, Spatial heterogeneity and the design of immunization programs, *Math. Biosci.* 72 (1984) 83.
- [25] R.S. Bernstein et al., Simulating the control of a heterosexual HIV epidemic in a severely affected East African city, *Interfaces* 28 (1998) 101.
- [26] H.W. Hethcote, Gonorrhea modeling: a comparison of control methods, *Math. Biosci.* 58 (1982) 93.
- [27] N.J. Robinson et al., Modelling the impact of alternative HIV intervention strategies in rural Uganda, *AIDS* 9 (1995) 1263.
- [28] I.M. Longini, E. Ackerman, L.R. Elveback, An optimization model for influenza A epidemics, *Math. Biosci.* 38 (1978) 141.
- [29] C. ReVelle, F. Feldmann, W. Lynn, An optimization model of tuberculosis epidemiology, *Manage. Sci.* 16 (1969) B190.
- [30] S.P. Sethi, Optimal quarantine programmes for controlling an epidemic spread, *J. Opnl. Res. Soc.* 29 (1978) 265.
- [31] D. Greenhalgh, Some results on optimal control applied to epidemics, *Math. Biosci.* 88 (1988) 125.
- [32] J. Muller, Optimal vaccination patterns in age-structured populations, *SIAM J. Appl. Math.* 59 (1998) 222.
- [33] D. Greenhalgh, Control of an epidemic spreading in a heterogeneously mixing population, *Math. Biosci.* 80 (1986) 23.
- [34] S.P. Sethi, P.W. Staats, Optimal control of some simple deterministic epidemic models, *J. Opnl. Res. Soc.* 29 (1978) 129.
- [35] S.P. Sethi, Quantitative guidelines for communicable disease control program: a complete synthesis, *Biometrics* 30 (1974) 681.
- [36] S. Blount, A. Galamboski, S. Yakowitz, Nonlinear and dynamic programming for epidemic intervention, *SIAM J. Appl. Math. Comput.* 86 (1997) 123.

- [37] K. Wickwire, Mathematical models for the control of pests and infectious diseases: a survey, *Theoret. Popul. Bio.* 11 (1977) 182.
- [38] V. Capasso, *Mathematical Structures of Epidemic Systems*, vol. 97, Springer, Berlin, 1993.
- [39] N.J. Bailey, in: *The Mathematical Theory of Infectious Diseases and Its Applications*, Hafner, New York, 1975.
- [40] C. Castillo-Chavez, S.A. Levin, C.A. Shoemaker, *Mathematical Approaches to Problems in Resource Management and Epidemiology*, vol. 81, Springer, Berlin, 1987.
- [41] R.J. Battjes, R.W. Pickens, Z. Amsel, HIV infection and AIDS risk behaviors among intravenous drug users entering methadone treatment in selected U.S. cities, *J. Acquir. Immune Defic. Syndr.* 4 (1991) 1148.
- [42] R.E. Booth, Gender differences in high-risk sex behaviors among heterosexual drug injectors and crack smokers, *Am. J. Drug Alcohol Abuse* 21 (1995) 419.
- [43] H.W. Hethcote, Qualitative analyses of communicable disease models, *Math. Biosci.* 28 (1976) 335.
- [44] S. Darke et al., Prevalence and predictors of psychopathology among opioid users, *Br. J. Addict.* 87 (1992) 771.
- [45] S. Darke, W. Swift, W. Hall, Prevalence, severity and correlates of psychological morbidity among methadone maintenance clients, *Addiction* 89 (1994) 211.
- [46] S. Darke et al., Predictors of injecting and injecting risk-taking behaviour among methadone-maintenance clients, *Addiction* 89 (1994) 311.
- [47] C. Contoreggi, V.E. Rexroad, W.R. Lange, Current management of infectious complications in the injecting drug user, *J. Subst. Abuse Treat.* 15 (1998) 95.
- [48] M. Prins et al., Pre-AIDS mortality from natural causes associated with HIV disease progression: evidence from the European Seroconverter Study among injecting drug users, *AIDS* 11 (1997) 1747.
- [49] J. von Limbeek et al., Prevalence of psychopathology in drug-addicted Dutch, *J. Subst. Abuse Treat.* 9 (1992) 43.
- [50] G.S. Zaric et al., The effect of protease inhibitors on the spread of HIV and the development of drug resistance: a simulation study, *Simulation* 71 (1998) 262.
- [51] A.P. Johnson-Masotti et al., Decision making in HIV prevention community planning: an integrative review, *J. Community Health* 25 (2000) 95.
- [52] P.E. Gill, W. Murray, M.H. Wright, in: *Practical Optimization*, Academic, New York, 1981.
- [53] D.G. Luenberger, in: *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1984.
- [54] R.M. Kaplan, J.W. Bush, Health-related quality of life measurement for evaluation research and policy analysis, *Health Psychology* 1 (1982) 61.
- [55] D.K. Owens, Interpretation of cost-effectiveness analyses, *J. Gen. Intern. Med.* 13 (1998) 716 (Editorial).
- [56] G.S. Zaric, M.L. Brandeau, *Dynamic resource allocation for epidemic control among multiple populations*, Technical Report, Department of Management Science and Engineering, Stanford University, Stanford, CA, 2000.
- [57] S.M. Adib et al., Relapse in sexual behavior among homosexual men: a 2-year follow-up from the Chicago MACS/CCS, *AIDS* 5 (1991) 757.
- [58] CDC, Patterns of sexual behavior change among homosexual/bisexual men – selected U.S. sites, 1987–1990, *MMWR* 40 (1991) 792.
- [59] N. El-Bassel, R.F. Schilling, 15-month followup of women methadone patients taught skills to reduce heterosexual HIV transmission, *Public Health Reports* 107 (1992) 500.
- [60] M.L. Elkstrand, T.J. Coates, Maintenance of safer sexual behavior and predictors of risky sex: the San Francisco Men's Health Study, *Am. J. Public Health* 80 (1990) 973.