Health Care Manag Sci
DOI 10.1007/s10729-013-9240-4

HIV epidemic control—a model for optimal allocation of prevention and treatment resources

Sabina S. Alistar · Elisa F. Long · Margaret L. Brandeau · Eduard J. Beck

Received: 11 November 2012 / Accepted: 18 April 2013

Abstract With 33 million people living with human immunodeficiency virus (HIV) worldwide and 2.7 million new infections occurring annually, additional HIV prevention and treatment efforts are urgently needed. However, available resources for HIV control are limited and must be used efficiently to minimize the future spread of the epidemic. We develop a model to determine the appropriate resource allocation between expanded HIV prevention and treatment services. We create an epidemic model that incorporates multiple key populations with different transmission modes, as well as production functions that relate investment in prevention and treatment programs to changes in transmission and treatment rates. The goal is to allocate resources to minimize R_0, the reproductive rate of infection. We first develop a single-population model and determine the optimal resource allocation between HIV prevention and treatment. We extend the analysis to multiple independent populations, with resource allocation among interventions and populations. We then include the effects of HIV transmission between key populations. We apply our model to examine HIV epidemic control in two different settings, Uganda and Russia. As part of these applications, we develop a novel approach for estimating empirical HIV program production functions. Our study provides insights into the important question of resource allocation for a country's optimal response to its HIV epidemic and provides a practical approach for decision makers. Better decisions about allocating limited HIV resources can improve response to the epidemic and increase access to HIV prevention and treatment services for millions of people worldwide.

Keywords HIV · Resource allocation · Basic reproduction number · Epidemic control

S. S. Alistar (✉) · M. L. Brandeau
Department of Management Science and Engineering, Stanford University, Stanford, CA, USA
e-mail: ssabina@stanford.edu

M. L. Brandeau
e-mail: brandeau@stanford.edu

E. F. Long
School of Management, Yale University, New Haven, CT, USA
e-mail: elisa.long@yale.edu

E. J. Beck
Programme Branch, UNAIDS, Geneva, Switzerland
e-mail: becke@unaids.org

1 Introduction

The HIV epidemic continues to be a major global health concern, with more than 33 million people infected worldwide [1]. Despite billions of dollars devoted to HIV prevention and treatment [2], 2.7 million people acquired HIV in 2010 [3]. The number of HIV-infected persons receiving antiretroviral therapy (ART) increased from 4 million in 2008 to 6.6 million in 2010, but almost half of eligible individuals are still not receiving life-saving treatment [3]. Moreover, for every person starting ART, an estimated two more become HIV-infected [4].

Many vulnerable populations lack access to proven HIV prevention programs. In 2009, only 8 % of injection drug users (IDUs) worldwide received drug substitution therapy and only 4 % of HIV-infected IDUs received ART [5]. In Eastern Europe and Central Asia, only 10 % of HIV prevention resources were provided to the highest risk populations, including IDUs, commercial sex workers, and men who have sex with men (MSM), despite the majority of new infections occurring among these groups [1]. In several sub-Saharan African countries, heterosexual contact and mother-to-child transmission account for most new infections. The remaining 30-40 % of new cases occur among IDUs, sex workers and clients, or MSM, but these groups collectively receive less than 2 % of all prevention resources [1].

Published online: 23 June 2013

To curb the epidemic, HIV prevention and treatment efforts must be further increased. However, available funds fall short of the estimated need [6], so efficient allocation of limited resources among HIV prevention and treatment services is crucial. Unfortunately, there is no "one size fits all" recipe for controlling HIV. Because the HIV epidemic varies across and within countries, the most effective allocation of resources will depend on local epidemic characteristics, program feasibility, and available funding [7–10].

Determining the optimal allocation of resources among competing HIV prevention and treatment programs is complex because the epidemic grows nonlinearly, interventions are not necessarily additive (health benefits of two interventions implemented simultaneously may not be equal to the sum of health benefits if implemented separately), and the relationship between resources spent and subsequent risk reduction is typically nonlinear (twice the investment in prevention may not achieve twice the risk reduction) [11–15]. Several studies have investigated resource allocation for HIV control, accounting for these nonlinear relationships (e.g., [16–19]), but such theoretical models may be difficult for decision makers to apply in practice. Some researchers have created numerical simulation models for HIV resource allocation (e.g., [8, 20–24]), but developing and implementing such simulations may be impractical for decision makers [25].

We develop a model to determine the optimal allocation of resources between HIV prevention and treatment, with the objective of minimizing the basic reproduction number, a measure of an epidemic's persistence in a population [26]. Our optimization model includes a simple epidemic model that governs the dynamics of HIV transmission (within and between multiple risk groups), disease progression, and death. We also empirically estimate production functions that relate investment in prevention and treatment programs to changes in epidemic model parameters.

We first develop a single-population model and determine the optimal mix of prevention and treatment (Section 2). We extend the analysis to multiple independent populations, with resource allocation among interventions and populations (Section 3). We then include the effects of cross-infection between populations (Section 4). We present a novel method for estimating HIV prevention program production functions, using a reduced-form structural estimation technique parameterized with data on past investments in prevention programs. As illustrative examples, we apply our model to examine HIV epidemic control in two settings, Uganda and Russia, which have different epidemics, key populations, and transmission modes (Section 5). We conclude with discussion of our findings (Section 6).

Our study offers several contributions. Our study is the first to link epidemiology to optimal allocation of resources between prevention and treatment for infectious disease control, while empirically estimating production functions that translate monetary investments in prevention to changes in transmission rates. Second, we systematically explore the implications of both treatment and prevention program scale up on the evolution of the epidemic, as quantified by the basic reproduction number. Third, we show in our examples for Russia and Uganda that, even with a simple estimation technique, insights can be gained about the future trajectory of the epidemic in specific settings.

2 Single population

2.1 SIT model

We construct a dynamic compartmental model with three disease states, susceptible, infected, and treated (SIT), for a single population (Fig. 1a). This model defines the epidemic's spread over time as individuals transition between compartments due to disease transmission, treatment, birth, or death, and is specified by the following differential equations:

$$\frac{dS}{dt} = bN - \beta(I + \gamma T)\frac{S}{N} - bS \tag{1}$$

$$\frac{dI}{dt} = \beta(I + \gamma T)\frac{S}{N} - \nu I - d_1 I \tag{2}$$

$$\frac{dT}{dt} = \nu I - d_2 T \tag{3}$$

$$N = S + I + T \tag{4}$$

where S, I, and T denote the number of susceptible, infected, and treated individuals at time t, respectively, and N denotes the total population size at time t. For compactness we omit the argument t. We assume that the law of mass action applies: new infections occur at a rate proportional to the size of the uninfected and infected compartments. The parameter β denotes the contact rate sufficient to allow transmission. We assume that both infected and treated individuals may infect the susceptible population, but that treatment reduces infectivity by a factor γ ($0 \leq \gamma \leq 1$). Infected individuals begin treatment at rate ν. Untreated infected individuals die at rate d_1 and treated individuals die at rate d_2. The non-disease-related mortality rate is b. In general, $d_1 > d_2 > b$. We assume that new individuals enter the susceptible compartment at rate bN.

2.2 Basic reproduction number

We compute the basic reproduction number (R_0), the average number of secondary infections caused by a typical

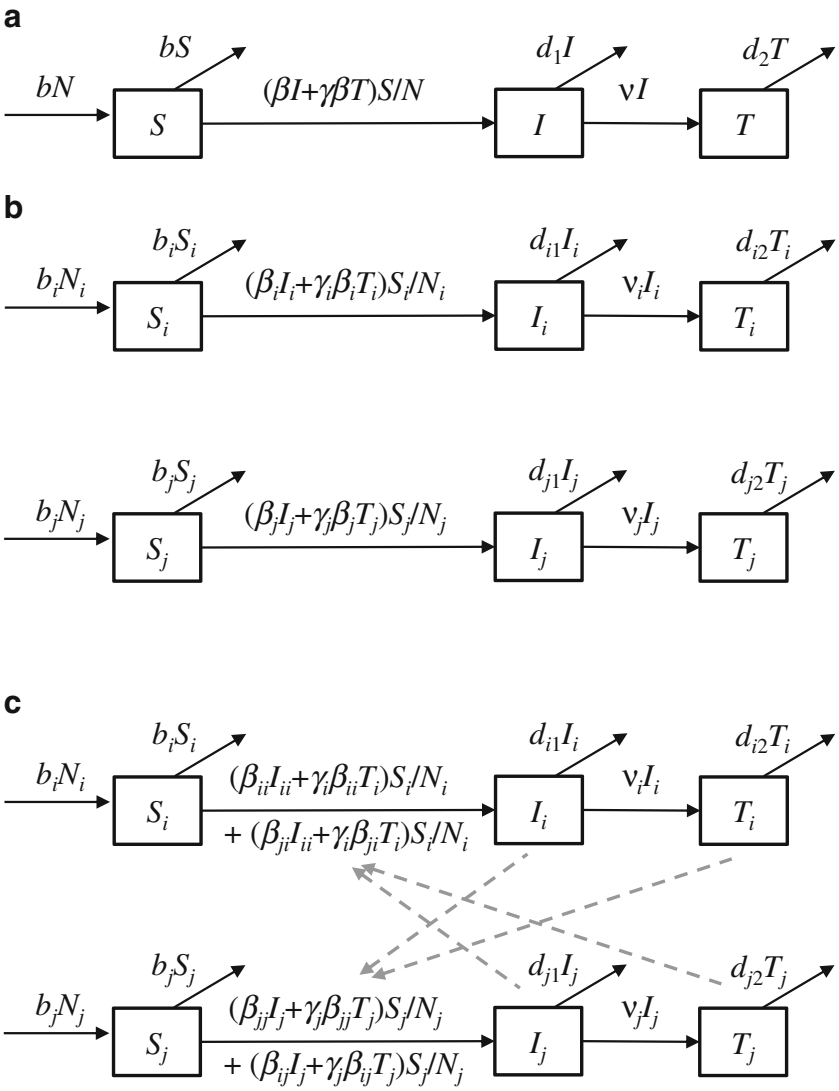

Fig. 1 Schematic diagrams of models. **a** Single population model. **b** Multiple independent population model. **c** Model with two interacting populations

infected individual in a susceptible population [27, 28] (see Appendix):

$$R_0 = \frac{\gamma\beta\nu + \beta d_2}{(\nu + d_1)d_2}. \tag{5}$$

R_0 has been widely used to evaluate the stability of epidemics and the conditions under which a disease remains endemic in a population [26]. If $R_0<1$, the disease-free equilibrium is asymptotically stable and the epidemic dies out, whereas if $R_0>1$, the epidemic persists. R_0 has been estimated for various diseases both from epidemiological data and from mathematical models of epidemics [28, 29], and has been particularly useful for modeling influenza [30, 31]. It is also useful for modeling sexually transmitted diseases such as HIV and evaluating the potential for disease-free equilibria [27]—a concept that is relevant to the stated goals of international HIV control programs to achieve "zero new HIV infections" [32] and "an AIDS-free generation" [33].

2.3 Effect of prevention and treatment programs

We develop a framework for analyzing the impact of treatment and prevention programs on R_0. An intervention aimed at improving disease *prevention* would reduce β, whereas an intervention aimed at increasing access to *treatment* would increase ν. The relationship between spending and the effects of these interventions may be nonlinear. For example, the marginal effectiveness of a prevention program may diminish as more individuals are reached. Hence, we incorporate *production functions* that translate monetary investment in a prevention or treatment program into changes in transmission and treatment rates.

Let x_1 and x_2 denote the investment in prevention and treatment, respectively. We define

$\beta(x_1)$ = contact rate sufficient for disease transmission if x_1 is spent on prevention.

$\nu(x_2)$ = treatment initiation rate if x_2 is spent on treatment.

We assume these functions are differentiable. A number of researchers have investigated production functions for HIV prevention programs [11–14, 18, 34–37]. Kaplan [14] estimated a production function for the New Haven Needle Exchange Program similar to the function $\beta(x_1)$ shown in Fig. 2a. The program had diminishing returns with investment, and the author estimated that there was a limit to total risk reduction, regardless of expenditure. A linear production function for treatment (Fig. 2b) may be plausible in many cases [20, 22, 38–43]. In Section 5, we propose a technique that uses data from real-world settings and a simplified (SIT) epidemic model to estimate such functions.

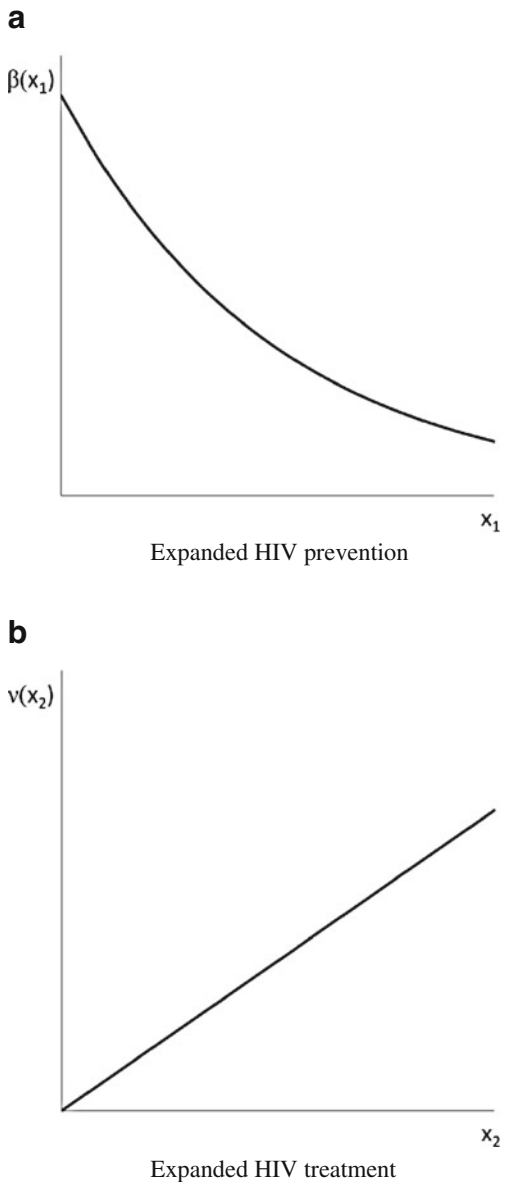

Fig. 2 Illustrative production functions for expanded HIV prevention (**a**) and treatment (**b**)

The basic reproduction number can now be written as a function of expenditure on prevention and treatment:

$$R_0(x_1, x_2) = \frac{\gamma\beta(x_1)\nu(x_2) + \beta(x_1)d_2}{[\nu(x_2) + d_1]d_2}. \tag{6}$$

2.4 Optimization model

To determine the optimal allocation of funds between HIV prevention and treatment, we construct an optimization model to minimize $R_0(x_1, x_2)$, subject to an available budget C:

$$\begin{aligned} &\text{minimize } R_0(x_1, x_2) \\ &\text{subject to } x_1 + x_2 \le C \\ &x_1, x_2 \ge 0 \end{aligned} \tag{7}$$

We now establish properties of the optimal solution. All proofs are in the Appendix. We impose the following two assumptions:

Assumption 1: $\beta(x_1)$ is monotonically decreasing.
Assumption 2: $\nu(x_2)$ is monotonically increasing and $\gamma < \frac{d_2}{d_1}$.

If either assumption holds, then the optimal investment (x_1^*, x_2^*) will utilize the entire budget, C. This means that if additional investment in prevention (Assumption 1) or treatment (Assumption 2) reduces the net infectivity of the population, then the optimal allocation should scale up prevention or treatment, as expected. The condition that $\beta(x_1)$ is monotonically decreasing implies that additional investment in prevention reduces overall transmission in the population. A monotonically increasing function $\nu(x_2)$ means that increased spending on treatment leads to an increase in the treatment initiation rate. The condition $\gamma<d_2/d_1$ implies that the advantages of treatment outweigh the disadvantages, in terms of total population infectivity. If $d_1>d_2$, then treatment reduces mortality and increases life expectancy. However, treated individuals are also less infectious by a factor of γ. If infectivity is sufficiently reduced to offset the increased life expectancy ($\gamma<d_2/d_1$) and $\nu(x_2)$ is monotonically increasing then it is optimal to utilize the entire budget.

Proposition 1 If $\beta(x_1)$ is decreasing and convex and $\nu(x_2)$ is increasing and linear, then $R_0(x_1, x_2)$ is quasi-convex.

Proposition 2 If $(\bar{x}_1, \bar{x}_2)$ satisfies the Karush-Kuhn-Tucker conditions (A2)-(A8) and if $\nabla R_0(\bar{x}_1, \bar{x}_2) \neq 0$, then the conditions are sufficient for optimality and $(\bar{x}_1, \bar{x}_2)$ is a global minimum.

Propositions 1 and 2 provide necessary and sufficient conditions for optimality. We now characterize the conditions under which it is optimal to invest in prevention or treatment alone, or a combination of the two. For the

remainder of the paper, we assume $\beta(x_1)$ is decreasing and convex and $\nu(x_2)$ is increasing and linear.

2.5 Optimal resource allocation

Proposition 3 It is optimal to invest only in prevention or treatment under the following conditions:

Case 1: Investing only in prevention $\left(x_1^* = C, x_2^* = 0\right)$ is optimal if $\gamma > \frac{d_2}{d_1}$, or if $\gamma < \frac{d_2}{d_1}$ and

$$\frac{d}{dx_1}\beta(C) < \frac{\beta(C)\frac{d}{dx_2}\nu(0)[\gamma d_1 - d_2]}{[\nu(0) + d_1][\gamma\nu(0) + d_2].}$$

Case 2: Investing only in treatment $\left(x_1^* = 0, x_2^* = C\right)$ is optimal if $\gamma < \frac{d_2}{d_1}$ and

$$\frac{d}{dx_1}\beta(0) > \frac{\beta(0)\frac{d}{dx_2}\nu(C)[\gamma d_1 - d_2]}{[\nu(C) + d_1][\gamma\nu(C) + d_2].}$$

Since R_0 is quasiconvex, these conditions are necessary and sufficient for optimality. If $\gamma > d_2/d_1$ (Case 1), the reduction in infectivity due to treatment does not offset the additional infections occurring from the increased life expectancy, and investing only in prevention is optimal. If $\gamma < d_2/d_1$ (Case 2), then it may be optimal to invest in treatment. In this case, if investing more funds in treatment leads to a greater reduction in R_0 than investing solely in prevention, then it is optimal to invest in treatment only. Note that minimizing R_0 accounts for new HIV infections but ignores the benefits of increased life expectancy and improved quality of life for infected individuals receiving treatment. In sensitivity analysis we consider the objective of maximizing quality-adjusted life years (QALYs) gained.

Proposition 4 It is optimal to invest in both prevention and treatment $\left(x_1^* = x^*, x_2^* = C - x^*\right)$ under the following condition:

$$\frac{d}{dx}\beta(x)\Big|_{x=x^*} = \frac{\beta(x)|_{x=x^*}\frac{d}{dx}\nu(x)\big|_{x=C-x^*}[\gamma d_1 - d_2]}{[\nu(C - x^*) + d_1][\gamma\nu(C - x^*) + d_2].}$$

Corollary 1 Let $\nu(x_2)=kx_2$. If $\frac{d}{dx_1}\beta(0) < -\beta(0)/\left(C + \frac{d_1}{k}\right)$, then it is optimal to invest some amount in prevention $\left(x_1^* > 0, x_2^* \geq 0\right)$.

Corollary 1 defines the condition where the marginal gain from increasing investment in prevention above zero is greater than the marginal loss from decreasing investment in treatment.

Corollary 2 Let $\nu(x_2)=kx_2$ and $\frac{d}{dx_1}\beta(0) < -1/\left(C + \frac{d_1}{k}\right)$. Suppose $\gamma=0$ and suppose the optimal investment is $\left(x_1^* > 0, x_2^* \geq 0\right)$. For all values of $\widehat{\gamma} > 0$, then the optimal investment is $\left(\widehat{x}_1^* \geq x_1^*, \widehat{x}_2^* \leq x_2^*\right)$.

Corollary 2 says that if investment in prevention is part of the optimal intervention package even when treatment completely eliminates infectivity ($\gamma=0$), then prevention will be part of the optimal solution for settings when treatment is less effective ($\widehat{\gamma} > 0$).

Table 1 shows the optimal allocation strategies and derived analytical conditions for optimality. It is straightforward to develop examples using reasonable parameter values and varying production functions in which it is optimal to allocate all of the budget to prevention, or to treatment, or to a combination of the two. Examples are shown in Fig. 3 (details in Appendix).

We have assumed a fixed budget, C. As the available budget increases, the investment decision will change with the production functions $\beta(x_1)$ and $\nu(x_2)$, and the value of γ. Thus, it is important to use a framework that incorporates production functions and their impact on R_0. With an increasing or decreasing budget, and similar assumptions for $\beta(x_1)$ and $\nu(x_2)$, it is straightforward to show that all the available funds will be spent and to derive similar conditions for the new optimal investment by comparing the marginal return per dollar invested for each program.

3 Multiple non-interacting populations

We extend our analysis to the case of n non-interacting populations ($n \geq 2$) (Fig. 1b). Analogous to the single-

Table 1 Optimal allocation strategies and associated conditions

Allocation strategy	Optimal allocation	Conditions
All prevention	$\left(x_1^* = C, x_2^* = 0\right)$	$\gamma > \frac{d_2}{d_1}$; or $\gamma < \frac{d_2}{d_1}$ and $\frac{d}{dx_1}\beta(C) < \frac{\frac{d}{dx_2}\nu(0)[\gamma d_1 - d_2]}{[\nu(0)+d_1][\gamma\nu(0)+d_2]}$
All treatment	$\left(x_1^* = 0, x_2^* = C\right)$	$\gamma < \frac{d_2}{d_1}$ and $\frac{d}{dx_1}\beta(0) > \frac{\frac{d}{dx_2}\nu(C)[\gamma d_1 - d_2]}{[\nu(C)+d_1][\gamma\nu(C)+d_2]}$
Some prevention	$\left(x_1^* > 0, x_2^* \geq 0\right)$	$\frac{d}{dx_1}\beta(0) < -\frac{1}{C+\frac{d_1}{k}}$ if $\nu(x_2) = kx_2$

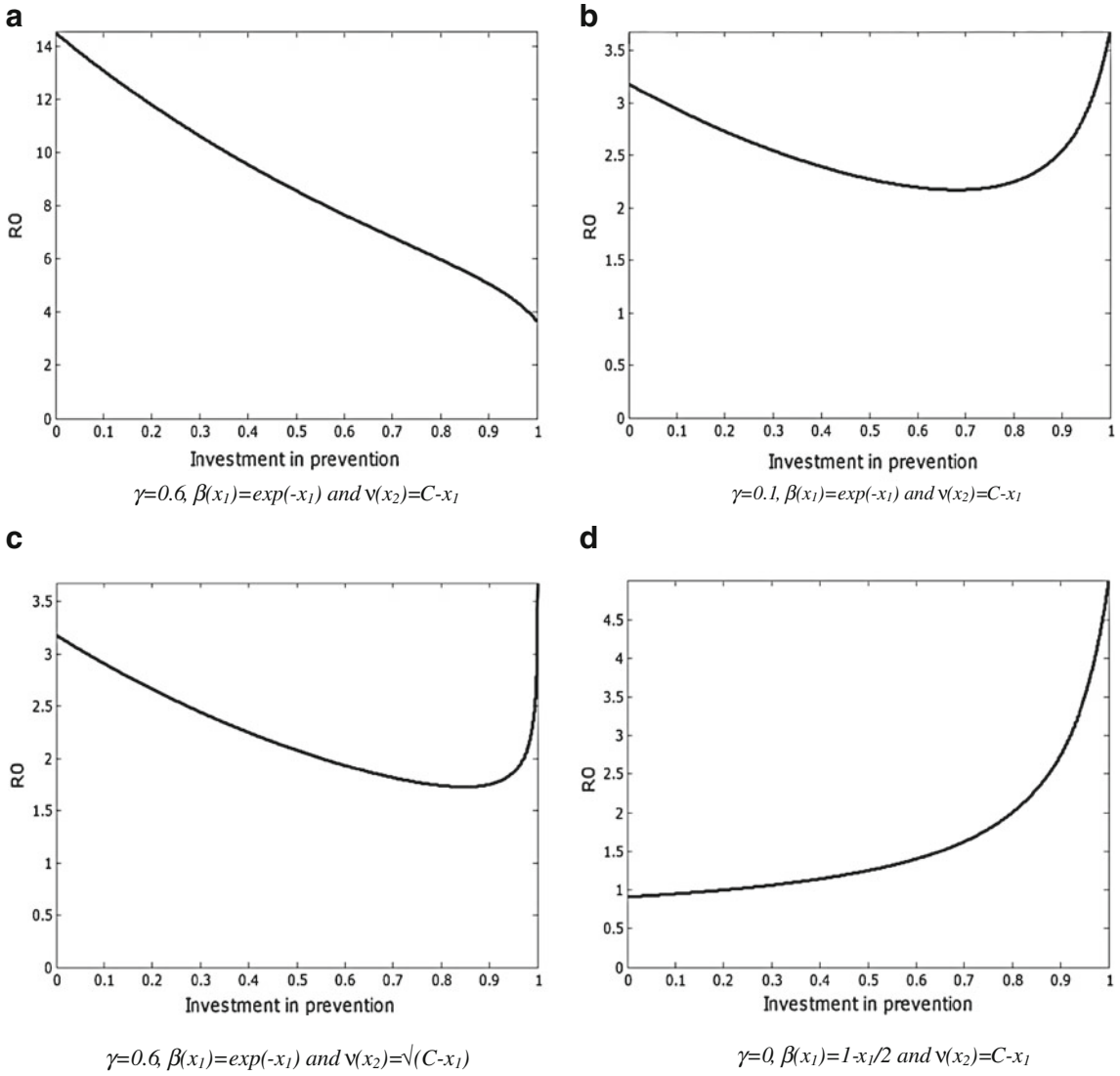

Fig. 3 Single-population model: $R_0(x_1, C - x_1)$ for various production function choices, assuming a budget $C=1$

population case, for each population $i=1,\ldots, n$ we have the following system of differential equations:

$$\frac{dS_i}{dt} = b_i N_i - (\beta_i I_i + \gamma_i \beta_i T_i) S_i / N_i - b_i S_i \tag{8}$$

$$\frac{dI_i}{dt} = (\beta_i I_i + \gamma_i \beta_i T_i) S_i / N_i - \nu_i I_i - d_{i1} I_i \tag{9}$$

$$\frac{dT_i}{dt} = \nu_i I_i - d_{i2} T_i \tag{10}$$

$$N_i = S_i + I_i + T_i \tag{11}$$

The basic reproduction number R_0 for this model is the sum of R_0 for each group i, which we denote by R_{0i}, and scale by weights α_i in our objective function to reflect decision maker priorities [28]:

$$R_0 = \sum_{i=1}^{n} \alpha_i R_{0i} = \sum_{i=1}^{n} \alpha_i \frac{\gamma_i \beta_i \nu_i + \beta_i d_{i2}}{(\nu_i + d_{i1}) d_{i2}}. \tag{12}$$

The constants α_i could reflect considerations such as population size, HIV incidence or prevalence levels, equity preferences, etc.

Analogous to the single-population case, R_{0i} is a function of prevention and treatment investments in population i:

$$R_{0i}(x_{i1}, x_{i2}) = \frac{\gamma_i \beta_i(x_{i1}) \nu_i(x_{i2}) + \beta_i(x_{i1}) d_{i2}}{[\nu_i(x_{i2}) + d_{i1}] d_{i2}}. \tag{13}$$

The optimization problem is now:

$$\begin{aligned} &\text{minimize } R_0 = \sum_{i=1}^{n} \alpha_i R_{0i}(x_{i1}, x_{i2}) \\ &\text{subject to } \sum_{i=1}^{n} x_{i1} + x_{i2} \le C \\ &x_{i1}, x_{i2} \ge 0, i = 1, \ldots, n \end{aligned} \tag{14}$$

This is equivalent to finding the optimal allocation (C_1, C_2,…, C_n) among the populations, where each C_i is a decision variable, and the optimal investment in treatment for each population ($x_{i2}=x_i$), with the remaining budget for each group (C_i-x_i) invested in prevention:

$$\begin{aligned} &\text{minimize} \sum_{i=1}^{n} \alpha_i R_{0i}(C_i, x_i) \\ &\text{subject to} \sum_{i=1}^{n} C_i \le C, \\ &C_i - x_i, x_i \ge 0, i = 1, \ldots, n \end{aligned} \tag{15}$$

For simplicity, we omit the subscript for investment in treatment ($x_{i2}=x_i$). We now develop results for two populations, denoted as populations i and j.

Proposition 5 Suppose at least one of the following conditions holds:

Condition 1: Either $\beta_i(C_i-x_i)$ or $\beta_j(C_j-x_j)$ is monotonically decreasing.

Condition 2: $v_i(x_i)$ is monotonically increasing and $\gamma_i < \frac{d_{j2}}{d_{j1}}$, or $v_j(x_j)$ is monotonically increasing and $\gamma_j < \frac{d_{j2}}{d_{j1}}$.

Then the optimal investment ($C_i-x_i, x_i, C_j-x_j, x_j$) will utilize the entire budget, C.

The proof of Proposition 5 shows that, under the given conditions, additional investment always reduces R_0, which means that increasing the budget available for epidemic control is worthwhile. Condition 1 implies that the transmission rate in either population i or j is decreasing as investment in prevention increases. Condition 2 implies that additional investment in treatment for population i or j increases the treatment entry rate, and that the reduction in infectivity outweighs the gain in life expectancy among those treated.

It is straightforward to show that for a given budget split (C_i, C_j), the optimal value of R_0 is obtained by optimizing R_{0i} and R_{0j} within the respective budget allocation. This finding is intuitive: the number of new infections occurring in each population are additive, and we are optimizing the use of funds within each population that receives any investment.

4 Two interacting populations

We now consider the case of two interacting populations i and j, where cross-infections can occur due to interactions between susceptible individuals in one population and infected individuals (treated or untreated) in the other population. This occurs in many real-world settings, where risk groups with different behaviors can transmit HIV among themselves and to other risk groups. A typical example is an epidemic driven by injection drug use. A small group of IDUs (comprising perhaps 1 % of the population) has very high HIV prevalence driven by needle sharing, but they can also transmit HIV to non-IDUs via sexual contact (e.g., [1, 20, 40]).

We use the same notation as before but now β_{ij} denotes the contact rate between an infected individual in population i and a susceptible individual in population j; we analogously define β_{ii}, β_{jj}, and β_{ji}. The model (Fig. 1c) is represented by the following equations:

$$\frac{dS_i}{dt} = b_i N_i - (\beta_{ii} I_i + \gamma_i \beta_{ii} T_i) S_i / N_i - (\beta_{ji} I_j + \gamma_j \beta_{ji} T_j) S_i / N_j - b_i S_i \tag{16}$$

$$\frac{dI_i}{dt} = (\beta_{ii} I_i + \gamma_i \beta_{ii} T_i) S_i / N_i + (\beta_{ji} I_j + \gamma_j \beta_{ji} T_j) S_i / N_j - v_i I_i - d_{i1} I_i \tag{17}$$

$$\frac{dT_i}{dt} = v_i I_i - d_{i2} T_i \tag{18}$$

$$N_i = S_i + I_i + T_i \tag{19}$$

$$\frac{dS_j}{dt} = b_j N_j - (\beta_{jj} I_j + \gamma_j \beta_{jj} T_j) S_j / N_j - (\beta_{ij} I_i + \gamma_i \beta_{ij} T_i) S_j / N_i - b_j S_j \tag{20}$$

$$\frac{dI_j}{dt} = (\beta_{jj} I_j + \gamma_j \beta_{jj} T_j) S_j / N_j + (\beta_{ij} I_i + \gamma_i \beta_{ij} T_i) S_j / N_i - v_j I_j - d_{j1} I_j \tag{21}$$

$$\frac{dT_j}{dt} = v_j I_j - d_{j2} T_j \tag{22}$$

$$N_j = S_j + I_j + T_j \tag{23}$$

The relative size of the two populations $r=N_i/N_j$ directly influences the epidemic's evolution. We compute R_0 for this system, which has not been previously

solved. In general, R_0 is the dominant eigenvalue of the linearized system of Eqs. (16)–(23) about the disease-free equilibrium (i.e., where $S_i = N_i$, for all i) [28]. To compute R_0, we use the next generation operator method. We compute the next generation matrix as follows (see Appendix for calculations):

$$FV^{-1} = \begin{bmatrix} \frac{\beta_{ii}(d_{i2}+\gamma_i\nu_i)}{d_{i2}(\nu_i+d_{i1})} & \frac{\gamma_i\beta_{ii}}{d_{i2}} & \frac{\beta_{ji}r(d_{j2}+\gamma_j\nu_j)}{d_{j2}(\nu_j+d_{j1})} & \frac{r\gamma_j\beta_{ji}}{d_{j2}} \\ 0 & 0 & 0 & 0 \\ \frac{\beta_{ij}(d_{i2}+\gamma_j\nu_i)}{d_{i2}(\nu_i+d_{i1})} & \frac{\gamma_j\beta_{ij}}{d_{i2}} & \frac{\beta_{ij}(d_{j2}+\gamma_i\nu_j)}{d_{j2}r(\nu_j+d_{j1})} & \frac{\beta_{ij}\gamma_i}{rd_{j2}} \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (24)$$

The (k,l) entry of the next generation matrix gives the expected number of new infections caused in compartment k by an infected individual in compartment l. R_0 is the spectral radius of this matrix and is given by the following formula:

$$R_0 = \frac{\beta_{ii}(d_{i2}+\gamma_i\nu_i)}{2d_{i2}(\nu_i+d_{i1})} + \frac{\beta_{ij}(d_{j2}+\gamma_i\nu_j)}{2d_{j2}r(\nu_j+d_{j1})} + \left(\left(\frac{\beta_{ii}(d_{i2}+\gamma_i\nu_i)}{2d_{i2}(\nu_i+d_{i1})} + \frac{\beta_{ij}(d_{j2}+\gamma_i\nu_j)}{2d_{j2}r(\nu_j+d_{j1})} \right)^2 - \frac{\beta_{ii}\beta_{ij}(d_{i2}+\gamma_i\nu_i)(d_{j2}+\gamma_i\nu_j)}{d_{i2}d_{j2}r(\nu_i+d_{i1})(\nu_j+d_{j1})} + \frac{\beta_{ij}\beta_{ji}r^2(d_{i2}+\gamma_j\nu_i)(d_{j2}+\gamma_j\nu_j)}{d_{i2}d_{j2}r(\nu_i+d_{i1})(\nu_j+d_{j1})} \right)^{1/2} \quad (25)$$

We again develop an optimization model to investigate the effects of prevention and treatment programs on epidemic outcomes. Analogous to the independent populations case, we assume that investment in prevention modifies β_{ii}, β_{ij}, β_{jj}, and β_{ji}, and investment in treatment modifies ν_i and ν_j . As before, we assume that x_{i1} and x_{i2} are the expenditures on prevention and treatment, respectively, in population i, with x_{j1} and x_{j2} defined analogously. We define production functions for prevention, $\beta_{ii}(x_{i1})$, $\beta_{ij}(x_{i1}, x_{j1})$, $\beta_{jj}(x_{j1})$, $\beta_{ji}(x_{i1}, x_{j1})$, and treatment, $\nu_i(x_{i2})$ and $\nu_j(x_{j2})$. Here we allow the cross-infection contact rates, β_{ij} and β_{ji}, to depend on spending in both populations. For example, investment in condoms for population i may reduce HIV transmission from infected persons in population i to susceptible persons in population j (β_{ij}), as well as transmission from infected persons in population j to susceptible persons in population i (β_{ji}).

Substituting the production functions into R_0 to obtain the term $R_0(x_{i1}, x_{i2}, x_{j1}, x_{j2})$, the resource allocation problem can be written as:

$$\begin{aligned} &\text{minimize } R_0(x_{i1}, x_{i2}, x_{j1}, x_{j2}) \\ &\text{subject to } x_{i1} + x_{i2} + x_{j1} + x_{j2} \le C \\ &x_{i1}, x_{i2}, x_{j1}, x_{j2} \ge 0 \end{aligned} \quad (26)$$

The expression $R_0(x_{i1}, x_{i2}, x_{j1}, x_{j2})$ is complex because of the nonlinearity of Eqs. (16)–(23). However, the optimal investment can be numerically calculated by solving the above optimization problem.

5 Model application: Uganda and Russia

We apply our resource allocation model using data for Uganda and Russia, countries with markedly different HIV epidemic characteristics. We estimate base values for the epidemic model parameters (Table 2) based on available epidemiological data. For each country, we estimate production functions based on available data and the type of epidemic. We illustrate how a simple estimation approach, combined with our nonlinear optimization model, can provide insights into the potential trade-offs between prevention and treatment in a given setting.

5.1 Uganda: a generalized epidemic

Uganda is often considered an HIV prevention success story. Adult HIV prevalence peaked at 15 % in 1990, and then declined throughout the 1990s and early 2000s due mostly to a national prevention campaign that involved

Table 2 Base case parameter values for epidemic models used in Uganda and Russia examples

Country	Compartment	Initial population size	Entry rate	Exit rate	Infectivity multiplier
Uganda	Susceptible (S)	7,895,000	0.05 (b)	0.05 (b)	0
	Infected (I)	1,285,000	0	0.15 (d_1)	1
	Treated (T)	0	0	0.10 (d_2)	0.1 (γ)
Russia	Susceptible IDU (S_1)	1,248,600	0.04 (b_1)	0.04 (b_1)	0
	Infected IDU (I_1)	165,400	0	0.18 (d_{11})	1
	Treated IDU (T_1)	0	0	0.13 (d_{12})	0.5 (γ_1)
	Susceptible non-IDU (S_2)	77,129,000	0.05 (b_2)	0.05 (b_2)	0
	Infected non-IDU (I_2)	7,400	0	0.15 (d_{12})	1
	Treated non-IDU (T_2)	0	0	0.10 (d_{22})	0.1 (γ_2)

extensive condom promotion [44]. In 2010, adult HIV prevalence was 6.5 %, indicating the presence of a *generalized epidemic*. An estimated 248,000 people received ART in 2010, amounting to 47 % of those in need [1]. With approximately 1.2 million people living with HIV and 100,000 new infections in 2010 out of an adult population of 17 million, additional prevention efforts are still needed [1]. A key public health question in Uganda is whether to allocate limited resources to additional prevention programs, invest in scaling up ART programs, or invest in both.

We assume that HIV prevention consists of condom promotion and associated counseling. We numerically estimate an exponentially decreasing production function for this program based on historic HIV epidemic data from Uganda. We assume a linear production function for treatment based on current ART costs. Using the single-population model and estimated production functions, we evaluate alternative prevention and treatment scale-up strategies for future years.

5.1.1 Production function estimates

We estimate Uganda's HIV prevention production function based on data for the number of condoms distributed and observed HIV prevalence levels from 1991 to 2000 (Table 3) [45]. We assume that the production function has the following reduced form (where x is a log-transform of the number of condoms distributed):

$$\beta(x) = \theta e^{-\lambda x} \tag{27}$$

We estimate the parameters θ and λ by applying the production function, $\beta(x)$, to our single-population SIT model and performing a least-squares minimization on observed and model-projected HIV prevalence. This is a novel method for estimating production functions in epidemiology. The term $\beta(x)$ varies from year to year, based on the number of condoms distributed. Importantly, we assume that no other prevention or treatment interventions are present during this time period, which is a reasonable assumption for Uganda. During the 1990s, public health officials focused predominantly on condom promotion, and virtually no individuals received ART. Settings where only one prevention program is in place are particularly well suited to our estimation technique. If more than one program is present, then additional observed HIV prevalence or incidence estimates are needed, as we discuss in the Russia example below.

Table 3 Condoms distributed and HIV prevalence in Uganda, 1991–2000

Year	Condoms distributed	Observed HIV prevalence
1991	0.5 million	14.0 %
1992	1.5 million	14.5 %
1993	2.0 million	14.0 %
1994	4.0 million	13.0 %
1995	6.5 million	12.0 %
1996	10.0 million	11.0 %
1997	16.0 million	10.2 %
1998	18.5 million	9.5 %
1999	16.0 million	8.5 %
2000	23.0 million	7.5 %

We instantiate Eqs. (1)–(4) with the following parameters for Uganda in 1991: $S(0)$=7,895,000; $I(0)$=1,285,000; $T(0)$=0; ν=0; γ=0.10; b=0.05; d_1=0.15; d_2=0.10. We estimated these parameter values based on available country statistics, country reports and published literature [1, 24, 44–46]. We numerically solve the system of equations and calculate projected HIV prevalence over time:

$$\text{Model prevalence}(t) = \frac{I(t) + T(t)}{S(t) + I(t) + T(t)} \tag{28}$$

To parametrize the production function, we perform a least-squares minimization using the following objective function:

$$\text{Min} \sum_t (\text{Observed prevalence}(t) - \text{Model prevalence}(t))^2 \tag{29}$$

We estimate the following production function for condoms distributed in Uganda, as shown in Fig. 4a:

$$\beta(x) = 6.412e^{-0.590x} \tag{30}$$

The coefficient of determination with this parametrization is 0.99. With this function, our model's projected HIV prevalence fits well with the observed decline in prevalence during the 1990s (Fig. 4b). We assume that the program costs $0.10 per condom [47, 48]. We assume a linear production function for treatment with a marginal annual cost of $500 per person [49].

5.1.2 Evaluation of alternative scale-up scenarios

We estimate that approximately 30 million condoms are used annually in Uganda, and that 248,000 people out of an estimated 530,000 with advanced HIV infection received ART in 2010. We compute R_0 (x_1, x_2) for our single-population SIT model (Eq. (6)), where x_1 and x_2 denote the investments in condoms and treatment, respectively. Under the status quo (x_1=0.10*30,000,000; x_2 =500*248,000), total program costs per year amount to $128 million, and R_0=

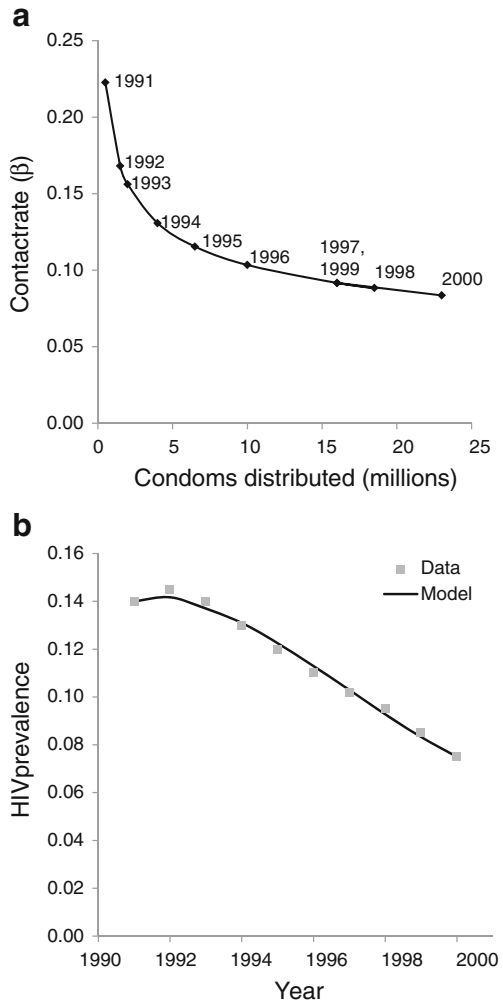

Fig. 4 Estimated production function and HIV prevalence in Uganda, 1991–2000. **a** Estimated condom program production function $\beta(x_1)$. **b** Actual and model-projected HIV prevalence

0.45, suggesting that the HIV epidemic in Uganda will decline in the long term (Table 4), but additional prevention efforts may be needed in the short term. In our analysis, we assume that resources cannot be transferred from the existing treatment program to condom distribution because it is unethical to cease treatment once it has been started.

As the budget increases from current levels, R_0 is minimized by allocating additional resources only to prevention (i.e., condom distribution). For example, if 100 million condoms are used annually, R_0=0.33, assuming constant treatment levels, for an additional cost of $7 million per year. Alternatively, a similar investment in scaling up treatment would provide ART to an additional 14,000 people per year, resulting in a value of R_0=0.42. To achieve an R_0= 0.33 using expanded treatment only, an additional 70,000 HIV-infected people must initiate ART at a cost of $35 million per year. In general, we find that investment in condom programs is preferred up to a value of 175 million condoms, beyond which the diminishing benefits of providing additional condoms suggest that treatment should then be prioritized. We also find that a combination strategy of expanded prevention and treatment can reduce R_0 to 0.17.

Using the estimated production functions and starting with 2010 demographic data, we also calculate new HIV infections until 2020 under various strategies assuming initial HIV prevalence of 6.5 % [1]. As shown in Table 4, the relative ranking of the strategies in terms of R_0 or HIV cases prevented is identical.

We use a nonlinear HIV prevention production function, which is a more conservative assumption than using a linear function. Despite this, we find that spending a modest amount of additional resources on increased condom distribution is a valuable investment. For all budgets considered, scaling up condoms to 175 million per year should be the priority investment if the objective is to reduce new infections. If resources permit, increasing treatment could also reduce the epidemic as well as morbidity and mortality among HIV-infected persons—health benefits not captured in this optimization framework. In sensitivity analysis (Section 5.3) we consider the alternative objective function of maximizing QALYs.

5.2 Russia: a concentrated epidemic

Russia is experiencing one of the fastest growing HIV epidemics in the world [1]. The first HIV outbreaks occurred in the late 1990s among IDUs, then spread rapidly via

Table 4 Uganda—costs and epidemic outcomes under alternative scale-up scenarios

Strategy	Condoms per year	People on treatment	Program costs per year	R_0	Reduced HIV cases 2011–2020
Status quo	30 million	248,000	$128 M	0.45	–
Condom increase	100 million	248,000	$135 M	0.33	31.5 %
Modest treatment increase	30 million	300,000	$143 M	0.39	17.0 %
Broad treatment increase	30 million	530,000	$268 M	0.23	59.1 %
Condom and treatment increase	100 million	530,000	$275 M	0.17	71.1 %

The estimated number of people in need of ART in Uganda in 2010 was 530,000 [1]

unsafe injection practices and needle sharing [50–52]. Russia currently has a *concentrated epidemic* [53], with most new HIV infections attributed to IDU-related transmission [54]. HIV prevalence among IDUs varies by region from 8 % to 64 %, with a mean of 37 % [55–57]. However, the increasing proportion of new infections (20–30 %) due to heterosexual transmission raises concerns about the risk of a generalized epidemic [57, 58]. HIV prevalence in adults is approximately 1.1 %, with more than 1 million individuals living with HIV [4, 58, 59].

Russia has recently scaled up its treatment and prevention efforts. In 2009 more than 71,000 individuals received ART, and 95 % of HIV-infected pregnant women were reached by programs to prevent mother-to-child transmission [4]. However, access to treatment remains low: only 15 % of eligible individuals received ART in 2007 [60], and more than 400,000 individuals will need treatment by 2015. Prevention efforts have lagged behind treatment: for every individual enrolled in ART, two more become newly HIV-infected [4]. Moreover, only 7 % of IDUs have access to harm reduction programs such as needle exchanges, well below the 60 % recommended by UNAIDS [61], and fewer than 1 % of IDUs receive ART [62].

Because the epidemic in Russia is concentrated among IDUs, we apply our model of two interacting populations (IDUs and non-IDUs). We assume that HIV prevention consists of harm reduction programs for IDUs and programs of education, awareness, and safe sex promotion for non-IDUs. Similar to our analysis for Uganda, we estimate exponentially decreasing prevention production functions based on historic HIV epidemic data, and assume linear ART production functions based on current costs. We use our epidemic model and the estimated production functions to compare alternative prevention and treatment scale-up strategies starting in 2007 and projected to 2020, considering scale up of prevention for IDUs and of treatment for both IDUs and non-IDUs.

5.2.1 Production function estimates

We estimate the production functions based on the number of harm reduction programs for IDUs in Russia, the number of HIV tests administered to pregnant women (as a proxy for prevention among non-IDUs), and estimated HIV prevalence (Table 5) [54, 57, 59, 60].

We assume that the production functions have the following forms (where x and y are log-transforms of the number of harm reduction programs and the number of administered HIV tests, respectively):

$$\beta_{11}(x) = \theta_{11} e^{-\lambda_{11} x} \tag{31}$$

$$\beta_{22}(y) = \theta_{22} e^{-\lambda_{22} y} \tag{32}$$

$$\beta_{12}(x) = \theta_{12} e^{-\lambda_{12} x} \tag{33}$$

$$\beta_{21}(y) = \theta_{21} e^{-\lambda_{21} y} \tag{34}$$

We denote by the index 1 parameters for IDUs, and by index 2 parameters for the general population. For simplicity, we assume that the sufficient contact rates for infections caused by IDUs ($\beta_{11}(x)$ and $\beta_{12}(x)$) depend solely on prevention efforts targeted to IDUs, and similarly $\beta_{22}(y)$ and $\beta_{21}(y)$ depend solely on prevention programs targeted to the general population.

We again estimate the parameters by performing a least-squares minimization on the observed and model-projected HIV prevalence for the two risk groups. We instantiate the model with 1999 data from Russia: $S_1(0)$=1,248,600; $I_1(0)$=165,400; $T_1(0)$=0; $S_2(0)$=77,129,000; $I_2(0)$=7,400; $T_2(0)$=0; $\gamma_1(0)$=0.50; $\gamma_2(0)$=0.10; b_1=0.04; b_2=0.05; d_{11}=0.18; d_{12}=0.13; d_{21}=0.15; d_{22}=0.10. We estimated these parameter values based on available country statistics, country reports and published literature [1, 51, 52, 54–58, 60, 63, 64]. Consistent with other analyses, we assume that

Table 5 Prevention efforts and HIV prevalence in Russia, 1999–2006

Year	Harm reduction programs	HIV tests for pregnant women	Estimated HIV prevalence in IDUs	Estimated HIV prevalence in non-IDUs
1999	24	2.6 million	11.7 %	0.010 %
2000	38	2.7 million	17.3 %	0.014 %
2001	44	2.7 million	19.3 %	0.155 %
2002	66	3.0 million	23.0 %	0.291 %
2003	74	3.3 million	27.0 %	0.422 %
2004	55	3.6 million	32.0 %	0.483 %
2005	46	3.9 million	33.0 %	0.566 %
2006	63	4.1 million	34.0 %	0.599 %

Russia had an estimated 2 million drug users in 2005 and 1.75 million pregnant women in 2007 [57]

treatment has a greater effect on transmission via sexual contacts than on transmission via needle sharing, since the latter is a more efficient mode of HIV transmission [20, 40]. We set $\nu_1=\nu_2=0$ initially but adjust this parameter in later years to reflect the initiation of ART scale up in 2003.

We estimate the following prevention program production functions for Russia:

$$\beta_{11}(x) = 2.726e^{-0.489x} \tag{35}$$

$$\beta_{22}(y) = 10.94e^{-0.323y} \tag{36}$$

$$\beta_{12}(x) = 0.059e^{-0.623x} \tag{37}$$

$$\beta_{21}(y) = 6.907e^{-0.573y} \tag{38}$$

Figure 5a shows as an example $\beta_{11}(x)$. Using these functions, our model's predicted HIV prevalence matches well the reported estimates for IDUs and non-IDUs (Fig. 5b). The coefficient of determination for both populations is 0.93. For the ART production function, we estimate a marginal cost per person of $6,500 based on reported data [57]. We assume a conservative cost per prevention program for IDUs of $100,000 [61]. This reflects the cost of providing one additional prevention center, which we assume will be able to reduce the rate of transmission according to the production function above. Fractional values may represent centers operating at less than full capacity.

5.2.2 Evaluation of alternative scale-up scenarios

In 2006, Russia's total HIV prevention budget was $62 million, and the HIV treatment budget was $97 million [57]. We estimated a 2007 status quo budget of $62 million for prevention and $173 million for treatment. We analyze alternative resource allocations to minimize R_0, assuming as before that existing programs would not be terminated (Table 6). Under the status quo, $R_0=2.08>1$, suggesting the epidemic will continue to grow. Implementing a 10 % increase in harm reduction programs for IDUs and scaling up to 30,000 people on treatment, we find that R_0 = 1.94 and new cases are reduced by 9 %. However, the optimal allocation of this $267 million budget is to invest all additional funds in

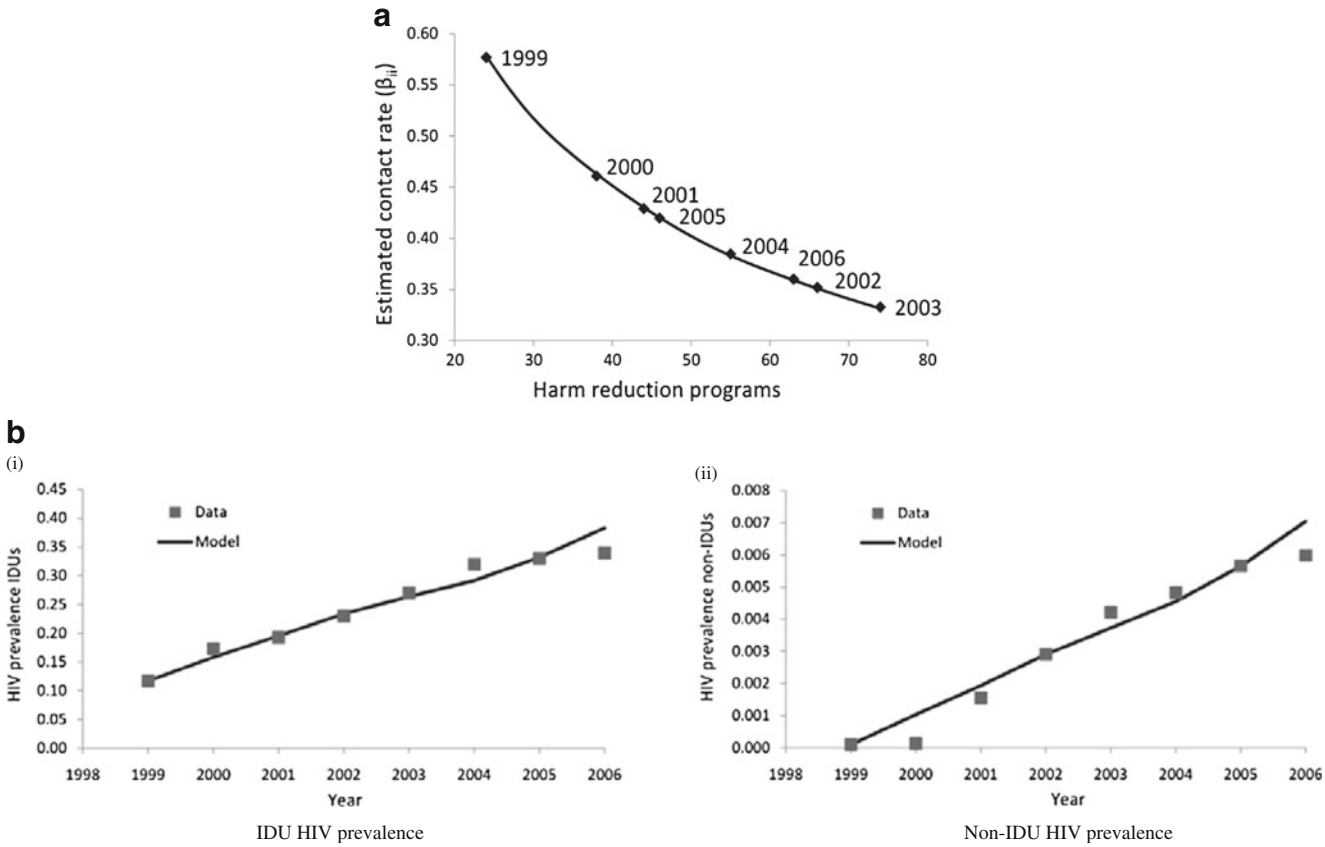

Fig. 5 Estimated production function and HIV prevalence in Russia, 1999–2006. **a** Estimated production function $\beta_{11}(x)$. **b** Actual and model-projected HIV prevalence in IDUs (i) and non-IDUs (ii)

Table 6 Russia—costs and epidemic outcomes under alternative scale-up scenarios of prevention for IDUs and treatment for IDUs and non-IDUs

Strategy	IDU prevention funding increase	People on treatment	Prevention costs per year	Treatment costs per year	R_0	Reduced HIV cases 2007–2020
Status quo	0 %	26,600	$62 M	$173 M	2.08	–
Increase prevention and treatment	10 %	30,000	$72 M	$195 M	1.94	9 %
Optimal allocation	506 %	26,600	$94 M	$173 M	0.82	66 %
Reduce R_0 below 1 (minimum budget)	306 %	26,600	$81.5 M	$173 M	1.00	58 %

The estimated number of people in need of ART in Russia in 2007 was 190,000 [60]

prevention for IDUs, thus reducing R_0 to 0.82 and new HIV cases by 66 %—a significant improvement over the planned allocation. The smallest total budget that would reduce R_0 below 1 is $254.5 million, with all of the incremental funds allocated to prevention for IDUs, reflecting their key role in the HIV epidemic in Russia. Once again, the relative ranking of strategies by R_0 or HIV cases averted is identical. If the total available budget is four times greater than we assumed ($1,100 M versus $267 M), then treating IDUs becomes part of the optimal solution.

Our analyses highlight the need to scale up prevention efforts targeted to IDUs to curb the HIV epidemic in Russia. Efforts directed solely to the non-IDU population have only a small impact on R_0 and HIV incidence. A comprehensive strategy that includes treatment scale up and increased IDU-targeted HIV prevention is the most efficient and effective option for controlling the HIV epidemic. These observations are independent of the absolute budget sizes, and reflect the relative effectiveness of each program at reducing new infections.

5.3 Sensitivity analysis

We now consider variations in key parameter values and modeling assumptions and compare their impact on resource allocation decisions for both Uganda and Russia.

Treatment costs have declined over time. If the cost of ART in Uganda falls below $400 per person (versus $500 in the base case), then scaling up ART (above the base case level) becomes part of the optimal solution. In Russia, if ART costs less than $1,350 per person (versus $6,500 in the base case), then scaling up ART becomes part of the optimal solution.

The extent to which ART reduces infectivity via needle sharing is uncertain. In our base case analysis for Russia, we assume γ_1=0.50. When we vary γ_1 between 0 (no transmission occurs if on treatment) and 0.90, we find that the optimal allocation is still to invest all incremental funds in harm reduction programs for IDUs.

Finally, because R_0 does not capture all the benefits of treatment, in sensitivity analysis we consider the alternative objective of maximizing QALYs gained. We assign quality adjustments for each health state: susceptible (1.0 for non-IDUs, 0.9 for IDUs), infected (0.8 for non-IDUs, 0.7 for IDUs), and treated (0.9 for non-IDUs, 0.8 for IDUs); due to the health consequences of drug use, IDUs have a lower quality of life than non-IDUs, and HIV treatment improves quality of life [42, 43, 65]. We also assign a terminal value at the end of the modeled time horizon to account for individuals who are still alive based on remaining life expectancy in each country and quality adjustment by state. We compute the total number of QALYs accumulated in the population by summing up the QALYs for each population group over the considered time horizon. We solve the model to maximize QALYs, and find that the optimal resource allocation for both Uganda and Russia is still to invest all additional funds in prevention programs. This occurs because with the current state of the epidemic in each country, preventing a new infection now offers a greater gain in QALYs than treating the infected individual later.

6 Discussion

This paper addresses a key public health question: what is the optimal investment in HIV prevention and treatment to maximize epidemic control? We have presented a mathematical method for determining optimal resource allocations, using the epidemiological parameter R_0 as an optimization criterion. The link between estimating epidemiological parameters, such as the reduction in disease transmission per marginal dollar spent, and optimal resource allocation has been insufficiently explored to date. We offer a straightforward approach for estimating HIV prevention "production functions" based on past program data. Application to two different epidemic settings suggests that our approach can provide qualitative insights into population-level effects of prevention and treatment on the HIV epidemic and the appropriate balance between them.

Our study illustrates a novel modeling approach and could be extended to include a more detailed epidemic structure. By deriving formulae for the basic reproduction number as a function of investment in prevention and treatment, we can determine necessary and sufficient conditions

for optimal investment in HIV prevention and treatment. Our model structure can be adapted to include key populations in a particular region, so the response can be tailored to the local epidemic. Additionally, our model for multiple interacting populations incorporates the transmission externality that one group imposes on the overall population, allowing for comparison of allocations across interventions and populations.

Using available data to estimate the marginal reduction in disease transmission as prevention or treatment investment increases (i.e., a production function) is essential for assessing the true impact of any public health intervention. The effectiveness of HIV prevention programs such as condom distribution or needle exchange often declines as a program is scaled up. Using illustrative production functions with identical budget constraints, we have shown that changes in the production function's shape can lead to different conclusions about the optimal allocation of resources among competing programs. In practice, the shape of the production functions can be determined from available data on program scale up (e.g., [11, 12, 35]) or mathematically from a few available data points (e.g., [14, 34]), but limited work has been done to non-parametrically estimate HIV production functions. A recent study employs linear production functions and a multiple-population SI model to optimize the use of HIV prevention resources in the U.S., but does not consider treatment and does not empirically estimate the parameterization for the production functions [66]. Estimating production functions requires time-series data, which may be unavailable in some settings. We assume a linear shape for the production function for treatment programs as this enables us to obtain analytical results. With non-linear treatment production functions, the function $R_0(x_1,x_2)$ is no longer quasi-convex. For our numerical examples, even with linear returns to scale for treatment, prevention is favored; a convex treatment production function, as implied by increasing marginal costs,would only favor prevention more. Further work is needed to collect and evaluate data on the costs and effectiveness of HIV prevention and treatment programs at different levels of implementation.

We make several simplifying assumptions to maintain tractability. We assume that treatment and prevention are independent, since in practice they are often implemented by different entities. However, treatment programs could lead to behavior change among those treated, and there may be economies of scope from implementing both types of programs in parallel. Our analysis could be extended to incorporate these effects by using production functions of the form $\beta(x_1,x_2)$. We use a simplified model of HIV, which ignores drug resistance, co-infection with other diseases, and complexities of sexual partnerships such as serial monogamy or concurrency. Our multiple interacting populations model could be adapted to reflect features such as male–female transmission asymmetry, or to include more disease stages to reflect changing infectivity over the course of HIV disease. We chose to keep the number of compartments small to gain analytical tractability and consequent insight; a more detailed model could be solved numerically.

Our optimization model minimizes the basic reproduction number, R_0, a theoretical epidemiologic concept most relevant in the early stages of an epidemic when a population is mostly susceptible. In many settings, HIV prevalence exceeds 10 % of the general population, and R_0 fails to fully capture these initial conditions. In our context, however, R_0 may be useful for showing the conditions under which the HIV epidemic can theoretically be eradicated, an important public health question. Minimizing R_0, similar to minimizing new HIV infections, reflects the full benefits of prevention and some of the benefits of treatment, but other objective functions such as maximizing life years or QALYs gained may be valid. Using R_0 as an objective function allows for closed-form solutions, which would not be available for other objective functions [17]. Policymakers may wish to add additional constraints, such as an equity constraint where each group receives some fraction of the available budget [67]. With a different objective function or constraints, numerical optimization could be performed. Many political, social, and ethical factors are relevant to HIV resource allocation [10, 25, 68] and these may not be captured by a single metric.

Our study provides insights into the important question of resource allocation for optimal scale up of HIV programs, and offers a useful approach for decision makers allocating resources between HIV prevention and treatment. In conjunction with UNAIDS, we are developing a more detailed resource allocation model, designed for practical use by decision makers [69]. Better decisions about allocating limited HIV resources, based on local data, can improve epidemic control and increase access to HIV prevention and treatment services for millions of people worldwide.

Acknowledgments This work was supported by Grant Number R01-DA15612 from the National Institute on Drug Abuse. Sabina Alistar was supported by a Stanford Graduate Fellowship.

Appendix

Proof of full budget spending The Lagrangian for the optimization problem in (7) is:

$$L(x_1,x_2,\lambda) = \frac{\lambda\beta(x_1)v(x_2)+\beta(x_1)d_2}{[v(x_2)+d_1]d_2} + \lambda(x_1+x_2-C) - \lambda_1 x_1 - \lambda_2 x_2. \tag{A1}$$

The following Karush-Kuhn-Tucker conditions must hold at optimality:

$$\frac{dL}{dx_1} = \frac{\beta'(x_1)[\gamma\nu(x_2) + d_2]}{[\nu(x_2) + d_1]d_2} + \lambda - \lambda_1 = 0 \tag{A2}$$

$$\frac{dL}{dx_2} = \frac{\beta(x_1)\nu'(x_2)[\gamma d_1 - d_2]}{[\nu(x_2) + d_1]^2 d_2} + \lambda - \lambda_2 = 0 \tag{A3}$$

$$\lambda\,(x_1 + x_2 - C) = 0 \tag{A4}$$

$$\lambda\, x_1 = 0 \tag{A5}$$

$$\lambda\, x_2 = 0 \tag{A6}$$

$$x_1 + x_2 - C \le 0 \tag{A7}$$

$$\lambda, \lambda_1, \lambda_2 \ge 0 \tag{A8}$$

If $\beta(x_1)$ is monotonically decreasing, then $\beta'(x_1)<0$ for all x_1. We know that $\lambda_1 \ge 0$. By Eq. (A2), $\lambda>0$, which implies that $x_1+x_2-C=0$. Similarly, if $\nu(x_2)$ is monotonically increasing, then $\nu'(x_2)>0$ for all x_2, and $\gamma < \frac{d_2}{d_1}$ is equivalent to $\gamma d_1 - d_2 < 0$. We know that $\lambda_2 \ge 0$. By Eq. (A3), $\lambda > 0$, which implies that $x_1 + x_2 - C = 0$. Hence, the optimal investment $\left(x_1^*, x_2^*\right)$ will utilize the entire budget, C.

Proof of proposition 1 If $\nu(x_2)$ is increasing and linear in x_2, then equivalently, $\nu(x_1)$ is decreasing and linear in x_1. Thus, we can rewrite R_0 in terms of one variable, x, assuming $x_2 = C - x_1$. $R_0(x) = \frac{\gamma\beta(x)\nu(x)+\beta(x)d_2}{[\nu(x)+d_1]d_2}$. Let $f(x)=\gamma\beta(x)\,\nu(x)$.

$$f'(x) = \gamma\beta(x)\nu'(x) + \beta'(x)\nu(x)] < 0$$
$$f''(x) = \gamma\beta(x)\nu''(x) + 2\beta'(x)\nu'(x) + \beta''(x)\nu(x)] > 0$$

because $\gamma>0$, $\beta(x)>0$, $\beta'(x)<0$, $\beta''(x)>0$, $\nu(x)>0$, $\nu'(x)<0$, $\nu''(x)=0$. Thus, $f(x)$ is decreasing and convex.

Let $g(x)=f(x)+\beta(x)d_2$. Then

$$g'(x) = f'(x) + d_2\beta'(x) < 0$$
$$g''(x) = f''(x) + d_2\beta''(x) > 0$$

because $d_2>0$, $f'(x)<0$, $f''(x)>0$, $\beta'(x)<0$, $\beta''(x)>0$. Thus, $g(x)$ is also decreasing and convex.

Let $h(y) = \frac{y}{\nu(x)}$, which is increasing in y. If $g(x)$ is convex and thus quasi-convex, then the composite function $h(g(x))$ is quasi-convex. Therefore, $R_0(x_1,x_2)$ is quasi-convex.

Proof of proposition 2 The proof follows from the Arrow-Enthoven sufficient conditions for optimality with quasi-convex objective functions [70].

Proof of proposition 3 We compare the marginal rate of substitution (MRS) for prevention $\left(\frac{dR_0}{dx_1}\right)$ and treatment $\left(\frac{dR_0}{dx_2}\right)$. As long as they are different, in order to find the optimal investment we will continue to shift the investment towards the program with the lower MRS, until we reach the budget limit C. At the optimal solution $\left(x_1^*, x_2^*\right)$, either the two MRS are equal $\left(\frac{dR_0}{dx_1} = \frac{dR_0}{dx_2}\right)$ so we are indifferent between the two programs, or we have already invested all the available budget in the most favorable intervention.

Case 1: Since we are minimizing, we invest only in prevention if

$$\frac{\frac{dR_0}{dx_1} < \frac{dR_0}{dx_2}}{\frac{\left[\frac{d}{dx_1}\beta(x_1)\right][\gamma\nu(x_2)+d_2]}{[\nu(x_2)+d_1]d_2} < \frac{\beta(x_1)\left[\frac{d}{dx_2}\nu(x_2)\right][\gamma d_1 - d_2]}{[\nu(x_2)+d_1]^2 d_2}} \tag{A9}$$

We have $\frac{d}{dx_1}\beta(x_1)$ because $\beta(x_1)$ is decreasing, and $\frac{d}{dx_2}\nu(x_2) > 0$ because $\nu(x_2)$ is increasing. Also, $\gamma\nu(x_2)+d_2>0$ and $\nu(x_2)+d_1>0$ because $\nu(x_2)>0$ and γ, d_1, and d_2 are positive constants. In order for (44) to always hold, $\gamma d_1 - d_2 > 0$, or equivalently, $\gamma > \frac{d_2}{d_1}$. If $\gamma d_1 - d_2 < 0$, or equivalently, $\gamma < \frac{d_2}{d_1}$, then the condition that must hold at $x_1=C$, $x_2=0$ is

$$\frac{d}{dx_1}\beta(x_1) < \frac{\beta(x_1)\frac{d}{dx_2}\nu(x_2)[\gamma d_1 - d_2]}{[\nu(x_2)+d_1][\gamma\nu(x_2)+d_2]}.$$

Case 2: We invest only in treatment if $\gamma < \frac{d_2}{d_1}$ and

$$\frac{\frac{dR_0}{dx_1} > \frac{dR_0}{dx_2}}{\frac{\left[\frac{d}{dx_1}\beta(x_1)\right][\gamma\nu(x_2)+d_2]}{[\nu(x_2)+d_1]d_2} > \frac{\beta(x_1)\left[\frac{d}{dx_2}\nu(x_2)\right][\gamma d_1 - d_2]}{[\nu(x_2)+d_1]^2 d_2}} \tag{A10}$$
$$\frac{d}{dx_1}\beta(x_1) > \frac{\beta(x_1)\frac{d}{dx_2}\nu(x_2)[\gamma d_1 - d_2]}{[\nu(x_2)+d_1][\gamma\nu(x_2)+d_2]}$$

which must hold at $x_1=0$, $x_2=C$.

Proof of proposition 4 We have shown that at optimal points the entire budget C is used, hence if $x_1^* = x^*$ then $x_2^* = C - x^*$.

$$\frac{\frac{dR_0}{dx_1} = \frac{dR_0}{dx_2}}{\frac{\left[\frac{d}{dx_1}\beta(x_1)\right][\gamma\nu(x_2)+d_2]}{[\nu(x_2)+d_1]d_2} = \frac{\beta(x_1)\left[\frac{d}{dx_2}\nu(x_2)\right][\gamma d_1 - d_2]}{[\nu(x_2)+d_1]^2 d_2}} \tag{A11}$$
$$\frac{d}{dx_1}\beta(x_1) = \frac{\beta(x_1)\frac{d}{dx_2}\nu(x_2)[\gamma d_1 - d_2]}{[\nu(x_2)+d_1][\gamma\nu(x_2)+d_2]}$$

For optimality, this condition must hold at $x_1^* = x^*$, $x_2^* = C - x^*$, which yields the condition described in the proposition statement.

Proof of corollary 1 Suppose γ=0 (that is, treatment completely eliminates the chance of disease transmission), which most optimistically favors treatment over prevention. By the previous proposition, we should invest only in treatment if at (x_1=0, x_2=C):

$$\begin{gathered}\frac{dR_0}{dx_1} > \frac{dR_0}{dx_2} \\ \frac{\left[\frac{d}{dx_1}\beta(x_1)\right][\gamma v(x_2)+d_2]}{[v(x_2)+d_1]d_2} > \frac{\beta(x_1)\left[\frac{d}{dx_2}v(x_2)\right][\gamma d_1 - d_2]}{[v(x_2)+d_1]^2 d_2} \\ \frac{d}{dx_1}\beta(x_1) > -\frac{d}{dx_2}v(x_2)\frac{1}{v(x_2)+d_1}\beta(x_1)\end{gathered}$$

because γ=0. If $v(x_2)=kx_2$, then at the point (x_1=0, x_2=C):

$$\begin{gathered}\frac{d}{dx_1}\beta(x_1) > -k\frac{1}{kx_2+d_1}\beta(0) \\ \frac{d}{dx_1}\beta(0) > -\frac{1}{C+\frac{d_1}{k}}\beta(0)\end{gathered} \tag{A12}$$

Equation (A12) defines the condition when it is optimal to spend only on treatment. Therefore, it is optimal to spend some amount on prevention ($x_1^* > 0, x_2^* \geq 0$) if:

$$\frac{d}{dx_1}\beta(0) < -\frac{\beta(0)}{C+\frac{d_1}{k}} \tag{A13}$$

Proof of corollary 2 By the envelope theorem, if the optimal investment at γ=0 includes some prevention, then as treatment becomes less effective at reducing infectivity ($\widehat{\gamma} > 0$), the optimal solution must not increase investment in treatment. Mathematically, as γ increases, $\frac{dR_0}{dx_2} = \frac{v'(x_2)[\gamma d_1 - d_2]}{[v(x_2)+d_1]^2 d_2}$ increases, so R_0 will increase as x_2 increases. Therefore, it cannot be optimal to have $\widehat{x}_2^* > x_2^*$.

Illustrative numerical examples of optimal allocation for different production functions and a single population Figure 3 shows numerical examples of R_0 for different production functions. These examples assume d_1=1/10 (average life expectancy of an infected individual is 10 years), d_2=1/25 (average life expectancy of a treated individual is 25 years). The budget, C=1, is entirely spent and the optimal allocation between prevention (x_1^*) and treatment (x_2^*) is given. In Fig. 3a, $\gamma > \frac{d_2}{d_1}$, and the reduction in infectivity due to treatment is too low to invest any amount in treatment ($x_1^* = 1, x_2^* = 0$). In Fig. 3b, $\gamma < \frac{d_2}{d_1}$ and $\frac{d}{dx_1}\beta(0) = -\infty$, and it is optimal to invest some amount in prevention ($x_1^* = 0.6821, x_2^* = 0.3179$). In Fig. 3c, γ and $\frac{d}{dx_1}\beta(0) = -\infty$, and it is optimal to allocate more to prevention because the treatment production function $v(x_2)$ has diminishing returns ($x_1^* = 0.8488, x_2^* = 0.1512$). In Fig. 3d, $\gamma < \frac{d_2}{d_1}$ and it is easy to verify that $\frac{d}{dx_1}\beta(0) > \frac{\frac{d}{dx_2}v(C)[\gamma d_1 - d_2]}{[v(C)+d_1][\gamma v(C)+d_2]}$. Thus, it is optimal to invest only in treatment ($x_1^* = 0, x_2^* = 1$).

Proof of proposition 5 The Lagrangian for the redefined problem is:

$$\begin{aligned}L(x_i, C_i, x_j, C_j, \lambda) &= \frac{\alpha_i\gamma_i\beta_i(C_i - x_i)v_i(x_i) + \beta_i(C_i - x_i)d_{i2}}{[v_i(x_i)+d_{i1}]d_{i2}} \\ &+ \frac{\alpha_j\gamma_j\beta_j(C_j - x_j)v_j(x_j) + \beta_j(C_j - x_j)d_{j2}}{[v_j(x_j)+d_{j1}]d_{j2}} \\ &+ \lambda(C_i + C_j - C) + \lambda_i(x_i - C_i) + \lambda_j(x_j - C_j).\end{aligned} \tag{A14}$$

The following Karush-Kuhn-Tucker conditions must hold at optimality:

$$\frac{dL}{dx_i} = \frac{-\alpha_i\beta_i'(C_i - x_i)[\gamma_i v_i(x_i) + d_{i2}]}{[v_i(x_i)+d_{i1}]d_{i2}} + \beta_i(C_i - x_i)\frac{\alpha_i v_i'(x_i)[\gamma_i d_{i1} - d_{i2}]}{[v_i(x_i)+d_{i1}]^2 d_{i2}} + \lambda_i = 0 \tag{A15}$$

$$\frac{dL}{dx_j} = \frac{-\alpha_j\beta_j'(C_j - x_j)[\gamma_j v_j(x_j) + d_{j2}]}{[v_j(x_j)+d_{j1}]d_{j2}} + \beta_j(C_j - x_j)\frac{\alpha_j v_j'(x_j)[\gamma_j d_{j1} - d_{j2}]}{[v_j(x_j)+d_{j1}]^2 d_{j2}} + \lambda_j = 0 \tag{A16}$$

$$\frac{dL}{dC_i} = \frac{\alpha_i\beta_i'(C_i - x_i)[\gamma_i v_i(x_i) + d_{12}]}{[v_i(x_i) + d_{i1}]d_{12}} + \lambda - \lambda_i = 0 \tag{A17}$$

$$\frac{dL}{dC_j} = \frac{\alpha_j\beta_j'(C_j - x_j)[\gamma_j v_j(x_j) + d_{j2}]}{[v_j(x_j) + d_{j1}]d_{j2}} + \lambda - \lambda_j = 0 \tag{A18}$$

$$\lambda(C_i + C_j - C) = 0 \tag{A19}$$

$$\lambda_i(x_i - C_i) = 0 \tag{A20}$$

$$\lambda_j(x_j - C_j) = 0 \tag{A21}$$

$$C_i + C_j - C \leq 0 \tag{A22}$$

$$x_i - C_i \leq 0 \tag{A23}$$

$$x_j - C_j \leq 0 \tag{A24}$$

$$\lambda, \lambda_i, \lambda_j \geq 0 \tag{A25}$$

Condition 1: If β_i $(C_i - x_i)$ is monotonically decreasing, then $\beta_i'(C_i - x_i)$ for all x_i. By Eq. (A17), $\lambda - \lambda_i > 0$, thus $\lambda > \lambda_i$. But $\lambda_i \geq 0$, hence $\lambda > 0$, which from Eq. (A19) implies $C_i + C_j - C = 0$. Similarly, if β_j $(C_j - x_j)$ is monotonically decreasing, then $\beta_j'(C_j - x_j)$ for all x_j. By (A18), $\lambda - \lambda_j > 0$, which implies that $C_i + C_j - C = 0$. If ν (x_2) is monotonically increasing, then $\nu'(x_2) > 0$ for all x_2, and γ is equivalent to $\gamma d_1 - d_2 < 0$. By (A3), $\lambda > 0$, which implies that $x_1 + x_2 - C = 0$. Hence, the optimal investment (x_1^*, x_2^*) will utilize the entire budget, C.

Condition 2: From (A15) and (A17), $\lambda = -\beta_i(C_i - x_i) \times \frac{\alpha_i \nu_i'(x_i)(\gamma_i d_{i1} - d_{i2})}{[\nu_i(x_i) + d_{i1}]^2 d_{i2}}$. If ν_i (x_i) is monotonically increasing and $\gamma_i < \frac{d_{i2}}{d_{i1}}$, then $\nu_i'(x_i) > 0$ and $\gamma_i d_{i1} - d_{i2} < 0$, hence $\lambda > 0$, which from (A19) implies $C_i + C_j - C = 0$. Similarly, from (A16) and (A18), $\lambda = -\beta_j(C_j - x_j) \frac{\alpha_j \nu_j'(x_j)(\gamma_j d_{j1} - d_{j2})}{[\nu_j(x_j) + d_{j1}]^2 d_{j2}}$. If ν_j (x_j) is monotonically increasing and $\gamma_j < \frac{d_{j2}}{d_{j1}}$, then $\nu_j'(x_j) > 0$ and $\gamma_i d_{i1} - d_{i2} < 0$, then $\lambda > 0$, which from (A19) implies $C_i + C_j - C = 0$.

Derivation of R_0 for one population For the given system, (1) – (4), we have:

$$
\begin{aligned}
z &= [I, T, S] \\
A(z) &= \left[\beta(I + \gamma T)\tfrac{S}{N}, 0, 0\right] \\
\Im(z) &= \left[\nu I + d_1 I, -\nu I + d_2 T, -bN + \beta(I + \gamma T)\tfrac{S}{N} + bS\right]
\end{aligned}
$$

The infected compartments are I and T, so the number of infected compartments is $m = 2$. Defining F and V as above, we obtain the following expression for the next generation matrix FV^{-1}:

$$
\begin{aligned}
F &= \begin{bmatrix} \beta & \beta\gamma \\ 0 & 0 \end{bmatrix}, V = \begin{bmatrix} \nu + d_1 & 0 \\ -\nu & d_2 \end{bmatrix} \\
FV^{-1} &= \begin{bmatrix} \frac{\gamma\beta\nu + \beta d_2}{(\nu + d_1)d_2} & \frac{\gamma\beta}{d_2} \\ 0 & 0 \end{bmatrix}
\end{aligned}
$$

The eigenvalues are $R_0 = \frac{\gamma\beta\nu + \beta d_2}{(\nu + d_1)d_2}$ and 0.

Derivation of R_0 for two interacting populations case We rewrite the differential equations for the two interacting populations, (16)–(23), to include the parameter $r = N_i/N_j$:

$$
\frac{dS_i}{dt} = b_i N_i - \left(\beta_{ii} I_i + \gamma_i \beta_{ii} T_i + r\beta_{ji} I_j + r\gamma_j \beta_{ji} T_j\right) S_i / N_i - b_i S_i \tag{A26}
$$

$$
\frac{dI_i}{dt} = \left(\beta_{ii} I_i + \gamma_i \beta_{ii} T_i + r\beta_{ji} I_j + r\gamma_j \beta_{ji} T_j\right) S_i / N_i - \nu_i I_i - d_{i1} I_i \tag{A27}
$$

$$
\frac{dT_i}{dt} = \nu_i I_i - d_{i2} T_i \tag{A28}
$$

$$
N_i = S_i + I_i + T_i \tag{A29}
$$

$$
\frac{dS_j}{dt} = b_j N_j - \left(\beta_{jj} I_j + \gamma_j \beta_{jj} T_j + 1/r * \beta_{ij} I_i + 1/r * \gamma_i \beta_{ij} T_i\right) S_j / N_j - b_j S_j \tag{A30}
$$

$$
\frac{dI_j}{dt} = \left(\beta_{jj} I_j + \gamma_j \beta_{jj} T_i + 1/r * \beta_{ij} I_i + 1/r * \gamma_i \beta_{ij} T_i\right) S_j / N_j - \nu_j I_j - d_{j1} I_j \tag{A31}
$$

$$
\frac{dT_j}{dt} = \nu_j I_j - d_{j2} T_j \tag{A32}
$$

$$
N_j = S_j + I_j + T_j \tag{A33}
$$

Let z be the vector representing the number of individuals in each compartment, $A_k(z)$ the rate of appearance of new infections in compartment z_k, and $\Im_k(z)$ the rate of change in compartment z_k by all other means. For the given system:

$$
\begin{aligned}
z &= \left[I_i, T_i, I_j, T_j, S_i, S_j\right] \\
A(z) &= \left[\left(\beta_{ii} I_i + \gamma_i \beta_{ii} T_i + r\beta_{ji} I_j + r\gamma_j \beta_{ji} T_j\right) S_i / N_i, 0,\right.
\end{aligned} \tag{A34}
$$

$$
\left.\left(\beta_{jj} I_j + \gamma_j \beta_{jj} T_j + 1/r * \beta_{ij} I_i + 1/r * \gamma_i \beta_{ij} T_i\right) S_j / N_j, 0, 0, 0\right] \tag{A35}
$$

$$
\begin{aligned}
\Im(z) = \big[&\nu_i I_i + d_{i1} I_i, -\nu_i I_i + d_{i2} T_i, \nu_j I_j + d_{j1} I_j, \nu_j I_j + d_{j2} T_j, \\
&-b_i N_i + \left(\beta_{ii} I_i + \gamma_i \beta_{ii} T_i + r\beta_{ji} I_j + r\gamma_j \beta_{ji} T_j\right) S_i / N_i + b_i S_i, \\
&-b_j N_j + \left(\beta_{jj} I_j + \gamma_j \beta_{jj} T_j + 1/r * \beta_{ij} I_i\right. \\
&\left. + 1/r * \gamma_i \beta_{ij} T_i\right) S_j / N_j + b_j S_j\big]
\end{aligned} \tag{A36}
$$

The infected compartments are I_i, T_i, I_j, and T_j, so the number of infected compartments is m=4. Defining F and V as below, we obtain the following expression for the next generation matrix FV^{-1}:

$$F = \left[\frac{\partial \mathcal{A}_k}{\partial z_l}(z_0)\right], V = \left[\frac{\partial \Im_k}{\partial z_l}(z_0)\right], \text{with} 1 \le k, l \le m$$

$$F = \begin{bmatrix} \beta_i & \gamma_i\beta_{ii} & r\beta_{ji} & r\gamma_j\beta_{ji} \\ 0 & 0 & 0 & 0 \\ \beta_j & \gamma_j\beta_{jj} & 1/r * \beta_{ij} & 1/r * \gamma_i\beta_{ij} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{A37}$$

$$V = \begin{bmatrix} \nu_i + d_{i1} & 0 & 0 & 0 \\ -\nu_i & d_{12} & 0 & 0 \\ 0 & 0 & \nu_j + d_{j1} & 0 \\ 0 & 0 & -\nu_j & d_{j2} \end{bmatrix}$$

$$FV^{-1} = \begin{bmatrix} \frac{\beta_{ii}(d_{i2}+\gamma_i\nu_i)}{d_{i2}(\nu_i+d_{i1})} & \frac{\gamma_i\beta_{ii}}{d_{i2}} & \frac{\beta_{ji}r(d_{j2}+\gamma_j\nu)}{d_{j2}(\nu_j+d_{j1})} & \frac{r\gamma_j\beta_{ji}}{d_{j2}} \\ 0 & 0 & 0 & 0 \\ \frac{\beta_{ij}(d_{i2}+\gamma_j\nu_i)}{d_{i2}(\nu_i+d_{i1})} & \frac{\gamma_j\beta_{jj}}{d_{i2}} & \frac{\beta_{ij}(d_{j2}+\gamma_i\nu_j)}{d_{j2}r(\nu_j+d_{j1})} & \frac{\beta_{ij}\gamma_i}{rd_{j2}} \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{A38}$$

The (k, l) entry of FV^{-1} gives the expected number of new infections caused in compartment k by an infected individual in compartment l. R_0 is defined as the spectral radius of FV^{-1} and is as given in Eq. (25).

References

1. Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO) (2010) UNAIDS report on the global AIDS epidemic 2010. http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf
2. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2009) What countries need: Investments needed for 2010 targets. http://data.unaids.org/pub/Report/2009/jc1681_what_countries_need_en.pdf
3. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2011) UNAIDS World AIDS Day report. How to get to zero: Faster. Smarter. Better. http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/jc2216_worldaidsday_report_2011_en.pdf
4. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2010) Outlook report 2010. http://data.unaids.org/pub/Report/2009/JC1796_Outlook_en.pdf
5. Mathers BM, Degenhardt L, Ali H, Wiessing L, Hickman M, Mattick RP, Myers B, Ambekar A, Strathdee SA (2009) Reference Group to the UN on HIV and Injecting Drug Use (2010) HIV prevention, treatment, and care services for people who inject drugs: a systematic review of global, regional, and national coverage. Lancet 375(9719):1014–1028
6. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2005) Resource needs for an expanded response to AIDS in low- and middle-income countries. http://data.unaids.org/pub/Report/2005/jc1255_resource_needs_en.pdf
7. Bertozzi SM, Laga M, Bautista-Arredondo S, Coutinho A (2008) Making HIV prevention programmes work. Lancet 372(9641):831–844
8. Hogan DR, Baltussen R, Hayashi C, Lauer JA, Salomon JA (2005) Cost effectiveness analysis of strategies to combat HIV/AIDS in developing countries. BMJ 331(7530):1431–1437
9. Horton R, Das P (2008) Putting prevention at the forefront of HIV/AIDS. Lancet 372(9637):421–422
10. Piot P, Bartos M, Larson H, Zewdie D, Mane P (2008) Coming to terms with complexity: a call to action for HIV prevention. Lancet 372(9641):845–859
11. Dandona L, Sisodia P, Kumar SP, Ramesh Y, Kumar AA, Rao MC, Marseille E, Someshwar M, Marshall N, Kahn J (2005) HIV prevention programmes for female sex workers in Andhra Pradesh, India: outputs, cost and efficiency. BMC Publ Health 5(1):98
12. Dandona L, Kumar SP, Ramesh Y, Rao MC, Kumar AA, Marseille E, Kahn JG, Dandona R (2008) Changing cost of HIV interventions in the context of scaling-up in India. AIDS 22(Suppl 1):S43–S49
13. Guinness L, Kumaranayake L, Hanson K (2007) A cost function for HIV prevention services: is there a 'u'—shape? Cost Eff Resour Alloc 5(1):13
14. Kaplan EH (1995) Economic analysis of needle exchange. AIDS 9(10):1113–1120
15. Kaplan EH (1998) Economic evaluation and HIV prevention community planning: a policy analyst's perspective. In: Holtgrave DR (ed) Handbook of economic evaluation of HIV prevention programs. Plenum Press
16. Brandeau ML, Zaric GS, Richter A (2003) Resource allocation for control of infectious diseases in multiple independent populations: beyond cost-effectiveness analysis. J Health Econ 22(4):575–598
17. Zaric GS, Brandeau ML (2001) Resource allocation for epidemic control over short time horizons. Math Biosci 171(1):33–58
18. Zaric GS, Brandeau ML (2001) Optimal investment in a portfolio of HIV prevention programs. Med Decis Making 21(5):391–408
19. Zaric GS, Brandeau ML (2002) Dynamic resource allocation for epidemic control in multiple populations. IMA J Math Appl Med Biol 19(4):235–255
20. Alistar SS, Owens DK, Brandeau ML (2011) Effectiveness and cost effectiveness of expanding harm reduction and antiretroviral therapy in a mixed HIV epidemic: a modeling analysis for Ukraine. PLoS Med 8(3):e1000423
21. Lasry A, Sansom SL, Hicks KA, Uzunangelov V (2012) Allocating HIV prevention funds in the United States: recommendations from an optimization model. PLoS One 7(6):e37545
22. Long EF, Brandeau ML, Owens DK (2010) The cost-effectiveness and population outcomes of expanded HIV screening and antiretroviral treatment in the United States. Ann Intern Med 153(12):778–789
23. Long EF, Stavert RR (2013) Portfolios of biomedical HIV interventions in South Africa: a cost-effectiveness analysis. J Gen Intern Med In press
24. Robinson NJ, Mulder DW, Auvert B, Hayes RJ (1995) Modelling the impact of alternative HIV intervention strategies in rural Uganda. AIDS 9(11):1263–1270
25. Alistar SS, Brandeau ML (2012) Decision making for HIV prevention and treatment scale up. Med Decis Making 32(1):105–117
26. Anderson RM, May RM (1991) Infectious diseases of humans: dynamics and control. Oxford University Press
27. Diekmann O, Heesterbeek JA, Metz JA (1990) On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J Math Biol 28(4):365–382
28. van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 180(1–2):29–48

29. Hethcote HW (2000) The mathematics of infectious diseases. SIAM Rev 42(4):599–653
30. Longini IM, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, Cummings DA, Halloran ME (2005) Containing pandemic influenza at the source. Science 309(5737):1083–1087
31. Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsirithaworn S, Burke DS (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437(7056):209–214
32. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2010) Getting to zero: 2011–2015 strategy. http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2010/JC2034_UNAIDS_Strategy_en.pdf
33. President's Emergency Fund for AIDS Relief (PEPFAR) (2012) PEPFAR blueprint: creating an AIDS-free generation. http://www.pepfar.gov/documents/organization/201386.pdf
34. Brandeau ML, Zaric GS, De Angelis V (2005) Improved allocation of HIV prevention resources: using information about prevention program production functions. Health Care Manag Sci 8(1):19–28
35. Marseille E, Dandona L, Marshall N, Gaist P, Bautista-Arredondo S, Rollins B, Bertozzi S, Coovadia J, Saba J, Lioznov D, Du Plessis J-A, Krupitsky E, Stanley N, Over M, Peryshkina A, Kumar SP, Muyingo S, Pitter C, Lundberg M, Kahn J (2007) HIV prevention costs and program scale: data from the PANCEA project in five low and middle-income countries. BMC Health Serv Res 7(1):108
36. Richter A, Brandeau ML, Owens DK (1999) An analysis of optimal resource allocation for prevention of infection with human immunodeficiency virus (HIV) in injection drug users and non-users. Med Decis Making 19(2):167–179
37. Brandeau ML, Zaric GS (2009) Optimal investment in HIV prevention programs: more is not always better. Health Care Manag Sci 12(1):27–37
38. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG (2009) Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet 373(9657):48–57
39. Lima VD, Johnston K, Hogg RS, Levy AR, Harrigan PR, Anema A, Montaner JS (2008) Expanded access to highly active antiretroviral therapy: a potentially powerful strategy to curb the growth of the HIV epidemic. J Infect Dis 198(1):59–67
40. Long EF, Brandeau ML, Galvin CM, Vinichenko T, Tole SP, Schwartz A, Sanders GD, Owens DK (2006) Effectiveness and cost-effectiveness of strategies to expand antiretroviral therapy in St. Petersburg, Russia. AIDS 20(17):2207–2215
41. Vijayaraghavan A, Efrusy MB, Mazonson PD, Ebrahim O, Sanne IM, Santas CC (2007) Cost-effectiveness of alternative strategies for initiating and monitoring highly active antiretroviral therapy in the developing world. J Acquir Immune Defic Syndr 46(1):91–100
42. Zaric GS, Barnett PG, Brandeau ML (2000) HIV transmission and the cost-effectiveness of methadone maintenance. Am J Public Health 90(7):1100–1111
43. Zaric GS, Brandeau ML, Barnett PG (2000) Methadone maintenance and HIV prevention: a cost-effectiveness analysis. Manag Sci 46(8):1013–1031
44. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2008) Epidemiological fact sheet on HIV and AIDS: Uganda 2008 update. http://apps.who.int/globalatlas/predefinedReports/EFS2008/full/EFS2008_UG.pdf
45. United States Agency for International Development (USAID) (2002) What happened in Uganda? Declining HIV prevalence, behavior change, and the national response. http://www.usaid.gov/our_work/global_health/aids/Countries/africa/uganda_report.pdf
46. Government of Uganda (2010) United Nations General Assembly Special Session (UNGASS) Country Progress Report, Uganda: January 2008–December 2009. http://data.unaids.org/pub/Report/2010/uganda_2010_country_progress_report_en.pdf
47. Altman LK (2005) U.S. blamed for condom shortage in fighting AIDS in Uganda. NY Times, Aug 30
48. Anonymous (2010) Uganda: Public irritated by yet another condom shortage. PlusNews, UN Office for the Coordination of Humanitarian Affairs. http://www.plusnews.org/Report.aspx?ReportId=89667
49. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2008) AIDS spending by categories and financing sources: Uganda. http://data.unaids.org/pub/report/2008/rt08_UGA_en.pdf
50. Hamers FF, Downs AM (2003) HIV in central and eastern Europe. Lancet 361(9362):1035–1044
51. Kelly JA, Amirkhanian YA (2003) The newest epidemic: a review of HIV/AIDS in Central and Eastern Europe. Int J STD AIDS 14(6):361–371
52. Twigg J (2007) HIV/AIDS in Russia: commitment, resources, momentum, challenges. http://csis.org/files/media/csis/pubs/071016_russiahivaids.pdf
53. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2007) Practical guidelines for intensifying HIV prevention: Towards universal access. http://data.unaids.org/pub/Manual/2007/20070306_prevention_guidelines_towards_universal_access_en.pdf
54. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2006) Country progress report of the Russian Federation on the implementation of the Declaration of Commitment on HIV/AIDS. Reporting period: January–December 2005. http://data.unaids.org/pub/Report/2006/2006_country_progress_report_russianfederation_en.pdf
55. Luo RF, Cofrancesco JJ (2006) Injection drug use and HIV transmission in Russia. AIDS 20(6):935–936
56. Mathers BM, Degenhardt L, Phillips B, Wiessing L, Hickman M, Strathdee SA, Wodak A, Panda S, Tyndall M, Toufik A, Mattick RP (2008) Global epidemiology of injecting drug use and HIV among people who inject drugs: a systematic review. Lancet 372(9651):1733–1745
57. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2008) Country progress report of the Russian Federation on the implementation of the Declaration of Commitment on HIV/AIDS. Reporting period: January 2006-December 2007. http://data.unaids.org/pub/Report/2008/russia_2008_country_progress_report_en.pdf
58. Burchell AN, Calzavara LM, Orekhovsky V, Ladnaya NN, for the Canada AIDS Russia Project and the Russian HIV Response Network (2008) Characterization of an emerging heterosexual HIV epidemic in Russia. Sex Transm Dis 35(9):807–813
59. World Health Organization (WHO) (2009) Towards universal access: scaling up priority HIV/AIDS interventions in the health sector. http://www.who.int/hiv/pub/tuapr_2009_en.pdf
60. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2008) Epidemiological fact sheet on HIV and AIDS: Russia 2008 update. http://apps.who.int/globalatlas/predefinedReports/EFS2008/full/EFS2008_RU.pdf
61. Open Health Institute (OHI) (2006) Harm reduction programs in the civilian and prison sectors of the Russian Federation: assessment of best practices. http://siteresources.worldbank.org/INTECAREGTOPHIVAIDS/Resources/HRBP_final.pdf
62. International Harm Reduction Association (IHRA) (2010) The global state of harm reduction 2010: key issues for broadening the response. http://www.ihra.net/files/2010/06/29/GlobalState2010_Web.pdf
63. Niccolai LM, Shcherbakova IS, Toussova OV, Kozlov AP, Heimer R (2009) The potential for bridging of HIV transmission in the Russian Federation: sex risk behaviors and HIV prevalence among drug users (DUs) and their non-DU sex partners. J Urban Health 85(Suppl 1):131–143

64. Niccolai LM, Toussova OV, Verevochkin SV, Barbour R, Heimer R, Kozlov AP (2010) High HIV prevalence, suboptimal HIV testing, and low knowledge of HIV-positive serostatus among injection drug users in St. Petersburg, Russia. AIDS Behav 14(4):932–941
65. Tengs TO, Lin TH (2002) A meta-analysis of utility estimates for HIV/AIDS. Med Decis Making 22(6):475–481
66. Lasry A, Sansom SL, Hicks KA, Uzunangelov V (2011) A model for allocating CDC's HIV prevention resources in the United States. Health Care Manage Sci 14(1):115–124
67. Kaplan EH, Merson MH (2002) Allocating HIV-preventionresources: balancing efficiency and equity. Am J Public Health 92(12):1905–1907
68. Institute of Medicine (IOM) (2001) No time to lose: Getting more from HIV prevention. http://home.mchsi.com/~apclc/8FedStudies1.pdf
69. Alistar SS, Brandeau ML, Beck EJ (2013) REACH: A practical HIV resource allocation tool for decision makers. In: Zaric GS (ed) Operations research and health care policy. Springer Publishers
70. Arrow KJ, Enthoven AC (1961) Quasi-concave programming. Econometrica 29(4):779–800