Published in final edited form as:

Med Decis Making. 2016 April ; 36(3): 391–409. doi:10.1177/0272989X15598528.

HIV Treatment and Prevention: A Simple Model to Determine Optimal Investment

Jessie L. Juusola, Ph.D. and Margaret L. Brandeau, Ph.D.¹

Department of Management Science and Engineering, Stanford University, Stanford, CA 94305

1. Introduction

In 2012, an estimated 35 million people were living with human immunodeficiency virus (HIV) worldwide and 2.3 million new infections occurred.¹ Both antiretroviral therapy (ART) and prevention are important HIV control measures. However, HIV control resources are limited. Striking the right balance between treatment and prevention can maximize health benefits in the most cost-effective manner.

ART for HIV-infected individuals is effective at both reducing infectivity and slowing disease progression.^{2–4} However, ART is expensive and requires access to medical care.⁵ Hence even in developed countries, a significant fraction of people eligible for ART do not receive it.^{6, 7} Moreover, it is estimated that worldwide, for every person newly enrolled in ART in 2010, two new infections occurred, indicating that more prevention efforts may be needed.⁶ Many HIV prevention methods are available and in use. Costs and effectiveness of the programs vary widely, from condom distribution, which may cost less than \$4 per person per year but have limited effectiveness,⁸ to daily preexposure prophylaxis (PrEP) with antiretroviral drugs, which may be highly effective at preventing disease acquisition but costs more than \$10,000 per person per year in the United States (US).^{9, 10}

Determining the appropriate investment of a limited budget in a portfolio of HIV treatment and prevention programs is complicated by several factors. First, the HIV epidemic grows nonlinearly, so preventing one infection today could lead to many infections averted in the future. Second, treatment and prevention programs may have overlapping benefits, since the same infection cannot be prevented twice. Individuals targeted by a prevention program may also reap the indirect benefits of a treatment program. Finally, programs may exhibit economies or diseconomies of scale as they are scaled up; thus, for example, twice the investment in prevention may not yield twice the reduction in risk.

This paper develops a simple framework for determining the optimal mix of treatment and prevention programs for HIV. The goal is to maximize health benefits, either quality-adjusted life years (QALYs) gained or HIV infections averted (an indirect measure of health benefits), subject to investment constraints. Most previous research on the cost-effectiveness of HIV control measures has focused on the cost-effectiveness of treatment or prevention

¹Corresponding author: Department of Management Science and Engineering, Huang Engineering Center, 475 Via Ortega, Stanford University, Stanford, CA 94305, brandeau@stanford.edu, Phone: (650)-725-1623, Fax: (650)-723-1614.

only,^{11–13} and studies of HIV resource allocation decisions have often focused only on HIV prevention programs.^{14–18} We consider investment in both. Prior studies that have directly compared treatment and prevention have used simulation models to assess program effectiveness,^{19–26} often ignoring their costs. We aim to provide a simple method of evaluating investment in treatment and prevention portfolios that is accessible to decision makers.

We simplify the resource allocation problem by approximating health benefits and costs with linear functions. Often, models that assume linearity of HIV intervention benefits do not capture epidemic effects or do so incompletely.^{11, 27–31} Our linear estimate of health effects accounts for both primary health benefits accruing to individuals who receive prevention or treatment as well as secondary health benefits accruing to other individuals who thereby avoid infection. Other studies have approximated epidemic effects similarly;^{32, 33} we expand on this by using the estimates as inputs to the resource allocation problem. Additionally, unlike many prior resource allocation models, we allow for subadditivity of program benefits and diseconomies of scale in program costs.

In Section 2 we formulate and solve our model, starting with a linear formulation and then expanding it to include overlap of treatment and prevention benefits and diseconomies of scale. In Section 3 we present methods for estimating model parameters. We illustrate our results in Section 4 with the examples of ART and two prevention programs – PrEP and community-based education – in men who have sex with men (MSM) in the US. We conclude with discussion in Section 5.

2. Model Formulation

We consider a population of adults including both HIV-infected and uninfected individuals. We assume the status quo includes baseline levels of treatment and prevention. We consider the decision of allocating funds between scaling up treatment for eligible infected individuals and implementing incremental prevention programs for uninfected individuals over a period of T years ($n-1$ prevention programs are available). We define x_1 as the number of incremental people starting ART due to the scale up and x_i ($i = 2, \dots, n$) as the number of uninfected people reached by prevention program i over T years. We denote overall program implementation by $\mathbf{x} = (x_1, x_2, \dots, x_n)$. We assume that programs are implemented early in the time horizon, and individuals remain on ART for their lifetime and receive prevention for T years. We assume that there is a minimum acceptable level of investment in each program, $x_i^{\min}$, and a maximum feasible level of program implementation, $x_i^{\max}$, thus, $0 \leq x_i^{\min} \leq x_i^{\max}$ ($i=1, \dots, n$). Table 1 shows all model notation.

We seek to maximize net present QALYs gained, $Q(\mathbf{x})$, or HIV infections averted, $IA(\mathbf{x})$, from the implementation of $\mathbf{x}$. These terms can be written as

$$Q(\mathbf{x}) = \int_0^T e^{-rt} [(QALYs \text{ experienced at time } t | \mathbf{x}) - (QALYs \text{ experienced at time } t | \mathbf{0})] dt$$

and

$$IA(\mathbf{x}) = \int_0^T e^{-rt} [(\text{New infections at time } t | \mathbf{0}) - (\text{New infections at time } t | \mathbf{x})] dt$$

where $\mathbf{0}$ is no incremental investment in any programs and r is the discount factor. QALYs gained is a standardized health outcome measure which is particularly appropriate for comparing treatment and prevention interventions³⁴ but many decision makers also consider HIV infections averted^{14–17, 35}, so we also include that objective.

2.1 Linear Resource Allocation Problem

We define a_1 as the estimated net present number of QALYs gained per person starting ART and a_i ($i = 2, \dots, n$) as the net present number of QALYs gained per person reached by prevention program i . We assume that these values include an estimate of the direct QALY benefit to each person reached by the programs plus an estimate of the indirect QALY benefit for persons who avoid infection as a result. We similarly define b_1 and b_i as the estimated net present number of (direct and indirect) infections averted per person starting ART and per person reached by prevention program i , respectively. We describe in Section 3 how these parameters can be estimated from empirical data.

Assuming no overlap in program benefits (that is, health benefits from simultaneous program implementation equal the sum of health benefits from each program implemented alone), we can write the following objective functions:

$$Q(\mathbf{x}) \cong \sum_{i=1}^n a_i x_i$$

$$IA(\mathbf{x}) \cong \sum_{i=1}^n b_i x_i$$

We constrain our objective functions with a budget B such that $C(\mathbf{x}) = B$, where $C(\mathbf{x})$ denotes the incremental costs in the modeled population with the implementation of $\mathbf{x}$. Specifically, we consider the incremental costs of treatment and prevention programs, discounted to the present at annual rate r ; this accounts for the cost of the intervention itself and the savings in medical care from improved health and reduced disease transmission.

In the simplest form of the problem, we model costs linearly. We define C_1 as the net present cost of starting one additional person on ART (for life) and C_i ($i = 2, \dots, n$) as the net present cost of reaching one uninfected person with prevention program i for T years.

Then total cost can be written as $C(\mathbf{x}) = \sum_{i=1}^n C_i x_i$. We assume that all C_i are positive, since a program that is costless or cost-saving would always be implemented at the maximum feasible level.

The resource allocation problem can be written as

P1

$$\max_x Q(\mathbf{x}) = \sum_{i=1}^n a_i x_i \text{ or } \max_x IA(\mathbf{x}) = \sum_{i=1}^n b_i x_i$$

$$s.t. \sum_{i=1}^n C_i x_i = B$$

$$x_i^{min} \leq x_i \leq x_i^{max}, i=1, \dots, n$$

P1 has a linear objective function and a linear budget constraint, so the optimal solution is a “greedy” solution (that is, funds are allocated sequentially to programs in increasing order of their cost-to-benefit ratio until all funds are spent).³⁶ Specifically, for maximizing QALYs, it is optimal to invest in x_i^{min} for all programs, and then invest any remaining budget as follows: invest up to x_i^{max} for the program with the smallest cost-to-benefit ratio, C_i/a_i (ties can be broken arbitrarily); if budget remains, repeat the process for the program with the next smallest C_i/a_i ; and continue until all budget is spent. This makes intuitive sense: it is optimal to invest in the programs that have the lowest ratio of costs to benefits. The results are similar when maximizing infections averted, but the comparison is between the ratios C_i/b_i .

We note that our problem is a resource allocation problem, not a cost-effectiveness analysis, so the cost-to-benefit ratios described here should not be confused with incremental cost-effectiveness ratios. We analyze the optimal allocation of resources among programs given a fixed budget constraint and assuming independence of intervention effects. (As we will describe in the following section, we capture potential dependence between program effects with an overlap multiplier (Problem P2 below).) Our analysis determines the optimal level of investment scale up in a given program, which is different from analyzing the cost-effectiveness of one program incremental to another; instead, the goal is to spend a fixed budget in the most effective manner possible. The cost-to-benefit ratios C_i/b_i are specific to the intervention considered, and do not depend on the other interventions under consideration.

2.2 Resource Allocation Problem with Overlap

Problem P1 assumes that benefits of treatment and prevention programs are purely additive. In practice the health benefits of simultaneously implemented programs can be less than the sum of the benefits of each program if implemented alone. For example, the number of HIV infections averted from a prevention program will be less when many eligible infected individuals receive ART than when few receive ART. This is because the average infectivity of infected individuals will be lower when many individuals receive ART, thus decreasing the baseline risk faced by uninfected individuals. A number of studies of concurrent prevention and treatment programs have observed this phenomenon.^{19, 25, 26, 31, 37}

We denote the magnitude of overlap in QALYs gained and infections averted by the implementation of $\mathbf{x}$ as $\delta_Q(\mathbf{x})$ and $\delta_{IA}(\mathbf{x})$, respectively. With this, we rewrite the resource allocation problem as:

P2

$$\max_{\mathbf{x}} Q(\mathbf{x}) = \sum_{i=1}^n a_i x_i - \delta_Q(\mathbf{x}) \text{ or } \max_{\mathbf{x}} IA(\mathbf{x}) = \sum_{i=1}^n b_i x_i - \delta_{IA}(\mathbf{x})$$

$$s.t. \sum_{i=1}^n C_i x_i = B$$

$$x_i^{min} \leq x_i \leq x_i^{max}, i=1, \dots, n$$

We assume that $\delta_Q(\mathbf{0}) = 0$ and $\delta_{IA}(\mathbf{0}) = 0$, and that these functions are increasing in x_i . We describe in Section 3 how these functions can be estimated. If $\delta_Q(\mathbf{x})$ and $\delta_{IA}(\mathbf{x})$ are concave in $\mathbf{x}$, then the objective function is convex. In this case, a boundary solution is optimal, and the same rules apply for optimally investing as in P1. If $\delta_Q(\mathbf{x})$ and $\delta_{IA}(\mathbf{x})$ are convex in $\mathbf{x}$, then the objective function is concave and the optimal solution can be determined from first-order conditions.³⁸ For other forms of $\delta_Q(\mathbf{x})$ and $\delta_{IA}(\mathbf{x})$, an analytical solution may be difficult to attain, but the problem can be solved numerically using standard methods.³⁸

One reasonable form for the overlap function is a Cobb-Douglas function. For example, for the case of ART and one prevention program, these functions are given by $\delta_Q(\mathbf{x}) = k_Q x_1 x_2$ and $\delta_{IA}(\mathbf{x}) = k_{IA} x_1 x_2$, where k_Q and k_{IA} are nonnegative constants. These functions account for increasing overlap as a function of each program's implementation, captured in the term $x_1 x_2$. If either x_1 or x_2 is zero, then $\delta_Q(\mathbf{x})$ and $\delta_{IA}(\mathbf{x})$ are zero, since overlap is only relevant when both programs are implemented. The Cobb-Douglas functions are neither convex nor concave. If these overlap functions are used, the problem can be solved numerically (see Appendix).

2.3 Resource Allocation Problem with Overlap and Increasing Marginal Costs

We further build on the resource allocation problem by considering increasing marginal costs with program scale up (that is, diseconomies of scale). For example, costs per person reached can increase with scale up if individuals reached first are the easiest (cheapest) to reach, or if individuals reached with scale up are less adherent and require more investment to achieve behavior change.^{37, 39} Increasing marginal costs can also serve as a proxy for diminishing returns in program effectiveness as a function of program scale up.⁴⁰ While secondary transmission benefits can lead to substantial numbers of infections averted at low program coverage, it is likely that the growth of health benefits slows as coverage increases and benefits approach their maximum level. This has been observed, for example, for PrEP for HIV prevention in MSM.¹⁰ As with overlap in health benefits, increasing marginal costs (or diminishing returns) may have little impact until program coverage is high.

We denote by $\lambda_i(x_i)$ the incremental average cost per person for treatment or prevention at program coverage x_i (above cost $C_i(x_i)$). We assume that $\lambda_i(x_i)$ is a nonnegative,

nondecreasing linear function,^{39, 41, 42} given by $\lambda_i(x_i) = \alpha_i x_i - \beta_i$. Thus, the incremental cost is incurred when coverage exceeds a threshold level β_i / α_i . The cost function is now

$$C_i(x_i) = \begin{cases} C_i x_i, & x_i^{\min} \leq x_i \leq \frac{\beta_i}{\alpha_i} \\ C_i x_i + (\alpha_i x_i - \beta_i) x_i, & \frac{\beta_i}{\alpha_i} \leq x_i \leq x_i^{\max} \end{cases}$$

where the terms C_i are as defined in the previous section.

With this modification, the resource allocation problem becomes

P3

$$\max_{\mathbf{x}} Q(\mathbf{x}) = \sum_{i=1}^n a_i x_i - \delta_Q \quad \text{or} \quad \max_{\mathbf{x}} IA(\mathbf{x}) = \sum_{i=1}^n b_i x_i - \delta_{IA}(\mathbf{x})$$

$$s.t. \sum_{i=1}^n C_i(x_i) = B$$

$$x_i^{\min} \leq x_i \leq x_i^{\max}, i=1, \dots, n$$

Since the budget constraint is piecewise, there are $n \times n$ cases to consider and, as before, the form of $\delta_Q(\mathbf{x})$ and $\delta_{IA}(\mathbf{x})$ affects the analytical tractability. The problem can be solved numerically, and some specific cases can be solved analytically.

We describe in the Appendix how to find the optimal solution to the resource allocation problems P2 and P3 for the special case of ART and one prevention program, with overlap given by Cobb-Douglas functions.

2.4 Evaluating Cost-Effectiveness

Models of HIV resource allocation typically assume a fixed budget to be allocated to maximize health impact,^{14–18, 35, 43} making the implicit assumption that any programs that are invested in will be incrementally cost-effective and that any minimum constraints on investment ($x_i^{\min}$ in our formulation) are appropriate. If there are no lower or upper limits on investment in any intervention, then the budget will always be spent and our framework will determine the set of investments that maximizes health benefit. In this case, there is a single dominant solution (same cost, equal or greater health benefit compared to all other possible investments) whose cost-effectiveness compared to the status quo can be readily determined.

When there are limits on investment in some or all interventions, then further analysis is needed to determine the efficiency and cost-effectiveness of alternative portfolios. In this case, one could solve the resource allocation problem separately for all possible combinations of interventions and then examine the cost-effectiveness plane to determine the incremental cost-effectiveness of various investments. We provide an example in Section 4.

3. Parameter Estimation

Since the model is intended to provide decision makers with a simple method of determining the optimal allocation of resources between treatment and prevention, it is important that planners can easily estimate model inputs. We now provide a set of simple rules for estimating key model parameters, summarized in Table 2. We define a discount factor

$$D = \sum_{t=1}^T \frac{1}{(1+r)^t}$$

that we use in calculating net present values of costs and health benefits.

3.1 Estimation of Health Benefits Parameters

QALY parameters (a_i)—The number of net present QALYs gained per person from ART, a_1 , can be calculated as the sum of the (net present) QALYs a treated individual gains from receiving ART, q_{tx} , and the (net present) QALYs gained from infections averted as a result of that individual receiving ART. The QALYs gained from infections averted depend on the net present QALY gain to an individual from avoiding infection, q_{IA} , and the number of infections averted, b_1 . Thus $a_1 \approx q_{tx} + q_{IA}(b_1/D)$. Estimates for q_{tx} and q_{IA} can be obtained from the available literature or estimated from existing disease models; we discuss how to estimate b_1 below.

The number of net present QALYs gained per person from prevention program i , a_i , is a function of both the individual's decreased chance of becoming infected and infections averted because of the epidemic effects from decreased incidence, b_i . We model the individual's decreased chance of becoming infected as $\sigma_T \epsilon_i$, where σ_T is the average chance a person acquires HIV in T years and ϵ_i is the efficacy of prevention intervention i in reducing the chance of HIV acquisition. This leads to $a_i \approx q_{IA}(\sigma_T \epsilon_i + (b_i/D))$. An estimate for ϵ_i can be obtained from the literature.

The method of estimating σ_T , the average chance that a person acquires HIV in T years, depends on the characteristics of the epidemic. For a stable epidemic, such as that in the overall population of MSM in the US, σ_T can be calculated based on the fairly stable yearly incidence. The chance of not acquiring HIV infection in one year can be calculated as the number of uninfected people at the end of the year divided by the number uninfected at the start of the year. The chance of acquiring HIV within T years can then be calculated as $1 - (\text{chance of not acquiring HIV in 1 year})^T$. For a rapidly growing epidemic, the calculation of σ_T must take into account the growth in incidence over time. In this case we can calculate σ_T using incidence estimates from two time points in the epidemic. This could be an estimate from past years' data, a current estimate, or a projection for the future, depending on what data is available to the decision maker. The overall σ_T can be estimated as the average of the σ_T for the low and high incidence time points, with each σ_T calculated using the methodology described above.

Infections averted parameters (b_i)—The net present number of infections averted per person from treatment or prevention, b_1 or b_i , also depends on the characteristics of the epidemic. For a stable epidemic, to estimate b_1 we first estimate the number of new infections each year caused by treated versus untreated infected individuals. We assume this does not change over time in a stable epidemic. We estimate the number of HIV-infected

individuals in the population who are currently receiving ART, based on available data. We then calculate the number of new infections caused by untreated individuals using the assumption that treated individuals' contribution to incidence is reduced by the efficacy of ART in reducing infectivity, ε_{tx} . The number of infections caused by an untreated individual in one year is then calculated as:

$$\frac{\text{\#new infections per year}}{\text{\#untreated} + (1 - \varepsilon_{tx}) \times \text{\#treated}}$$

and the number of infections caused by a treated individual in one year would be this reduced by a factor of $(1 - \varepsilon_{tx})$. We multiply these numbers by the discount factor D to estimate the net present number of new infections caused in T years. Then, the difference between the number caused by an untreated individual and the number caused by a treated individual is b_1 . For a rapidly growing epidemic, the methodology for estimating b_1 is similar, but we use incidence estimates from two time points, as we do for calculating σ_T . We calculate the number of infections caused by an untreated individual and a treated individual, as described above, in the low incidence year and in the high incidence year, and then use the average of those to calculate the net present number of new infections caused in T years by each type of individual. As before, the difference is b_1 .

We estimate b_i for prevention program i by multiplying an individual's reduction in the chance of acquiring infection ($\sigma_T \varepsilon_i$, as described above) by the net present number of infections he would cause over T years if he became infected. In a stable epidemic, the net present number of infections caused per infected person over T years is the number of annual new infections divided by the total number of infected people at the start of the year, multiplied by the discount factor:

$$\frac{\text{\#new infections per year}}{\text{\#infected at start of year}} \times D.$$

In a rapidly growing epidemic, we calculate the number of infections caused per person per year for two time points in the epidemic. To be conservative, since prevention programs may exhibit diminishing returns,^{10, 40, 44, 45} we use the smaller of the two estimates to calculate b_i .

3.2 Estimation of Cost Parameters (C_i)

We now provide simple rules for estimating the net present cost of starting (and maintaining) one additional person on ART, C_1 , and the net present cost of reaching one uninfected person with prevention program i for T years, C_i . Each parameter is a function of the cost of either ART or prevention for one person, as well as the savings from infections averted as a result of the person receiving ART or prevention. We define the net present savings from averting an HIV infection as s_{IA} . This is the difference in discounted lifetime healthcare costs for an infected individual and an uninfected individual; an estimate for s_{IA} can be obtained from the literature. For treatment, we assume that once an individual starts ART, he remains on treatment for the rest of his life. We estimate the net present lifetime

costs of treatment as c_1 . This is the difference in discounted lifetime healthcare costs for an infected individual receiving ART and an infected individual who never receives treatment; this parameter can be estimated from the literature. We then calculate C_1 as $C_1 \approx c_1 - s_{IA}(b_1/D)$.

Similarly, we define c_i as the discounted cost of prevention program i for one person for T years. This can either be obtained from the literature or calculated based on the costs of each component of the prevention program and the number of people reached by the program. In estimating C_i , we additionally factor in the savings in healthcare costs from a lower chance of each individual becoming HIV infected, $\sigma_T \epsilon_i$, and infections averted because of epidemic effects, b_i . This leads to $C_i \approx c_i - s_{IA}(\sigma_T \epsilon_i + (b_i/D))$.

3.3 Estimation of Overlap Functions ($\delta_Q(\mathbf{x})$ and $\delta_{IA}(\mathbf{x})$)

In the Appendix we outline methods for estimating k_Q and k_{IA} for the two-program problem with Cobb-Douglas production functions. These estimates are based on program effectiveness, average incidence, time horizon T , and initial size of the susceptible population. For the case of n programs, one approach to estimating overlap is to estimate the pairwise overlap between each pair of programs, using a method such as that described in the Appendix, and then include the corresponding overlap factors whenever investment in two or more programs is made. We illustrate this approach in Section 4.

In practice, overlap factors will generally be very small relative to the health benefits a_i and b_i and only noticeable when program implementation is high. Hence, it may be reasonable to ignore overlap when solving for the optimal treatment and prevention mix if implementation levels are relatively low. For the examples in Section 4, inclusion of overlap does not change the optimal solution.

4. Examples: Treatment and Prevention for MSM in the US

We illustrate our models with the two-program examples of ART and two types of prevention for MSM in the US: PrEP and community-based education (CBE), as well as a three-program example that considers all three interventions. MSM account for 56–61% of estimated new HIV infections in the US^{46, 47} and are an important group to reach with HIV control measures. Preexposure prophylaxis with daily antiretroviral drugs (ARVs) reduced HIV acquisition by 44% among MSM in the iPrEx study.⁹ However, PrEP is very costly, so it is important to examine investment in PrEP compared to investment in ART. Previous studies have assessed the cost-effectiveness of PrEP compared to no PrEP for MSM in the US but have not compared it with investment in ART.^{10, 27, 48} Community-based education, conversely, is a fairly low cost intervention aimed at reducing HIV risk behaviors. One such program, the Mpowerment Project, was found to modestly reduce the prevalence and frequency of unprotected anal intercourse in young MSM.⁴⁹ A previous study found the Mpowerment Project was cost-effective compared to no intervention, but did not compare it with investment in ART.⁵⁰

We estimated parameter values for ART, PrEP, and CBE (Table 3) using the guidelines described in Section 3. Where necessary, we drew values from the literature or estimated

them using a previously published model of the HIV epidemic in MSM in the US.¹⁰ We examined the general adult MSM population of approximately 4.3 million men.^{51, 52} We estimated 12.3% HIV prevalence in this population. Annual incidence is fairly stable at 0.8%.^{47, 51–54} We considered a time horizon of 20 years of program implementation and discounted costs, QALYs, and infections averted at 3% annually.³⁴ We considered a total budget of \$10 billion, estimated as 50% of the US federal budget for domestic HIV prevention.⁵⁵ We assumed $50,000 \leq x_1 \leq 400,000$ and $10,000 \leq x_i \leq 3,000,000$; these are lower and upper limits on the number of people receiving ART and prevention program i , respectively. We assumed that a non-zero minimum investment was required for prevention program i because various factors, such as ethical and societal considerations, are involved in setting investment priorities.

4.1 PrEP and ART

When considering the two-program problem of ART versus PrEP, for both P1 (which assumes linear costs and no intervention overlap) and P2 (which includes overlap in program effects), the optimal solution is to spend the minimum amount required on PrEP ($x_2=10,000$ people receiving PrEP) and invest the remaining budget in ART scale up ($x_1=72,786$ incremental people receiving ART). This is the case regardless of whether the objective is to maximize QALYs gained or infections averted. This is intuitive because the per person benefits from ART are greater than those from PrEP, and the per person costs are less. This allocation results in approximately 1,142,000 QALYs gained and 64,000 HIV infections averted over 20 years (Figure 1).

We also considered the case of diminishing returns for PrEP (problem P3). We estimated that the per person cost of PrEP increases for each person receiving PrEP once x_2 reaches 6,000. The costs of ART remain linear. However, because $x_2^{\min} > 6,000$, the marginal cost of PrEP is increasing. The optimal investment decision remains the same as for the case of linear costs of PrEP: implement $x_2^{\min}$ (10,000 people receiving PrEP) and invest the remaining budget in ART. Since implementing $x_2^{\min}$ costs more with nonlinear costs, it is only feasible to implement up to $x_1=72,769$ incremental people receiving ART. The numbers of QALYs gained and infections averted are just slightly lower than in P1 and P2 but within the rounding of our reporting.

4.2 Community-Based Education and ART

Community-based education is a much lower-cost intervention than PrEP, with a lower cost-to-health-benefit ratio than ART. In the two-program problem, for both P1 (linear costs and no intervention overlap) and P2 (which includes overlap in program effects), the optimal solution is to invest the maximum amount in CBE and invest the remaining budget in ART scale up. In this example, there is budget remaining after investing in $x_2^{\max}$ people reached by CBE, leading to the optimal solution of $x_1=77,754$ incremental people receiving ART and $x_2=3,000,000$ people reached by CBE. This is the case regardless of whether the objective is to maximize QALYs gained or infections averted. This allocation results in approximately 1,873,000 QALYs gained and 97,000 HIV infections averted over 20 years.

For a community-based intervention, costs may be higher than in the base case for certain cities and subgroups of MSM. Hence, we considered the case of increasing marginal costs for CBE (problem P3). We estimated that the per person cost of CBE increases once x_2 reaches 20,000. The costs of ART remain linear, but the costs of CBE are linear or quadratic, depending on the level of implementation. The QALY maximizing solution is $x_1=84,184$ incremental people receiving ART and $x_2=72,549$ people reached by CBE, with approximately 1.3 million QALYs gained. The solution that averts maximal infections is $x_1=84,446$ incremental people receiving ART and $x_2=48,397$ people reached by CBE, with approximately 74,000 HIV infections averted.

4.3 PrEP, Community-Based Education and ART

We also considered the three-program example of ART versus PrEP versus CBE. For P1, with the base case parameter values, the optimal solution is to spend the minimum amount required on PrEP ($x_2=10,000$ people receiving PrEP), the maximum possible on CBE ($x_3=3,000,000$ people receiving CBE) and invest the remaining budget in ART scale up ($x_1=65,877$ incremental people receiving ART). This is the case regardless of whether the objective is to maximize QALYs gained or infections averted. This results in approximately 1,701,000 QALYs gained and 88,000 HIV infections averted over 20 years. To apply problem framework P2, we approximate overlap in the three-program problem as the sum of the overlap between ART and PrEP and between ART and CBE. The optimal solution remains unchanged, but the resulting QALYs gained and infections averted are slightly lower due the effects of overlap (approximately 1,648,000 QALYs gained and approximately 83,000 HIV infections averted over 20 years). As is intuitive, the health outcomes when all three programs are considered fall between those attained in the two-program examples, but they are more similar to those attained by investment in ART and CBE, since most investment is allocated to CBE. When all three programs are considered, health benefits are diminished because of the constraint on minimum investment in PrEP, which is a less efficient intervention than CBE and ART; if there is no such constraint, no money is invested in PrEP and the optimal solution is the same as for the case of ART and CBE.

4.4 Comparison of Results with Dynamic Model Results

To assess the accuracy of our models, we compared the health benefits we estimated using P1 with those from a previously published dynamic model of HIV in MSM in the US.¹⁰ Figure 2 shows QALYs gained and infections averted as estimated by each method for varying levels of investment in ART, PrEP, and CBE. The results from the two models are fairly similar, suggesting that the assumption of linearity in our simple model is reasonable. The optimal solution of investing maximally in ART before investing in PrEP and investing in CBE before investing in ART is consistent between the two models. Where there are differences between the simple and dynamic models (for ART, the linear model accurately estimates QALYs but overestimates infections averted; for CBE and PrEP, the linear model accurately estimates infections averted but overestimates QALYs), the relative scale of ART, PrEP, and CBE benefits is still consistent between the two models, indicating that the simple model is likely “good enough” for informing resource allocation decisions.

4.5 Cost-Effectiveness Analysis

The above examples have considered various combinations of the three interventions and with constraints on levels of investment. Figure 3 shows net present costs and QALYs for the optimal portfolio corresponding to each possible combination of interventions, and assuming the upper and lower limits on investment as described above. We assumed that at most 3 million people could be reached by CBE, so when investing in just CBE, only \$816 million is spent. For all other sets of interventions, the entire \$10 billion budget is spent.

We can use the information in Figure 3 to ask the question: given the set of interventions considered in a portfolio, is the corresponding optimal investment cost-effective compared to the status quo? The least cost-effective portfolio occurs when only PrEP is considered (\$113,600/QALY gained compared to the status quo). The most cost-effective portfolio occurs CBE is considered (\$1200/QALY gained compared to the status quo). All other portfolios have incremental cost-effectiveness ratios (ICERs, compared to the status quo) that fall between these two values.

We can also ask the question: what combinations of interventions lie on the efficient frontier, and what is their ICER compared to the next least costly alternative on the efficient frontier? For our example, the portfolio of CBE only and the portfolio of CBE and ART lie on the efficient frontier. Adding ART to a portfolio that includes CBE costs \$7600/QALY gained. For this example, then, a decision maker could conclude that it is most cost-effective to invest just in CBE but is also cost-effective to invest in ART and CBE, and that portfolios that include PrEP or do not include CBE would be dominated by other alternatives.

4.6 Sensitivity Analysis

In sensitivity analysis, we assessed how several key factors – the efficacy of the prevention program, the time horizon of program implementation, and the nature of the epidemic in the modeled population – affect the results. Table 4 shows parameter values for the ART and PrEP scenarios examined. For simplicity, we assumed linear program costs.

In the iPrEx study, PrEP reduced HIV incidence by 73% in patients who reported or exhibited high adherence to the daily regimen.⁹ With this higher efficacy (versus 44% assumed in the base case), the number of QALYs gained and infections averted per person receiving PrEP increases, and the net present per person cost of PrEP decreases. However, ART is still a more efficient investment than PrEP and it is still optimal to spend the minimum amount on PrEP and invest the remaining budget in ART. Since the net per person cost of PrEP is lower, there is more budget remaining after investing in x_2^{min} so x_1 is higher than in the base case, with 73,316 incremental people receiving ART. Studies of PrEP among heterosexuals^{56–59} and injection drug users⁶⁰ have found reductions in risk of HIV acquisition ranging from 0% to 75%. For any of these efficacy levels, if achieved among MSM, our analysis shows that ART is a more efficient investment than PrEP.

The base case assumed that CBE reduces risky behavior by 8%. If CBE reduces risky behavior by only 4% or less, when the objective is to maximize infections averted, it becomes optimal to invest more than the minimum amount in ART scale up. The optimal

solution is to invest x_2^{min} in CBE and invest all remaining budget in ART scale up, for 84,616 incremental people receiving ART. When maximizing QALYs gained, it only becomes optimal to invest more than the minimum amount in ART scale up when CBE reduces risky behavior by less than 1%.

In the base case we considered a 20-year time horizon for PrEP use and ART scale up. Decision makers may prefer a shorter time horizon, since they are often concerned with short-term budgets rather than long-term expenditures. Accordingly, we examined a 5-year time horizon. Again, the optimal solution is to spend the minimum amount on PrEP and invest the remaining budget in ART scale up ($x_1=13,433$ incremental people receiving ART). However, shortening the time horizon leads to a greater decrease in the health benefits of PrEP, a_2 and b_2 , than in the benefits of ART, a_1 and b_1 (Figure 4). This is because treatment yields an immediate benefit, both in the QALY gain to the individual receiving treatment and in the protection conferred to others. The benefits of PrEP are more dependent on epidemic effects accumulating over time, while ART confers immediate protection to all partners of treated individuals.

The same trend can be seen with the comparison of ART and CBE. Shortening the time horizon to 5 years (versus 20 years in the base case) leads to a greater decrease in benefits from CBE than from ART. As before, the optimal solution is to invest only the minimum amount in ART scale up and to invest all remaining funds in CBE. However, in this case there is budget remaining after investing the maximum possible in CBE, leading to the optimal solution of $x_1=12,889$ incremental people receiving ART and $x_2=1,000,000$ reached by CBE.

The base case considered the general MSM population in the US, which has a fairly stable HIV epidemic. A rapidly growing epidemic presents very different conditions for disease control. In sensitivity analysis we considered an epidemic with prevalence growing from 5% to 18% over 20 years, and with incidence growing from 0.7% to 2.2% during this same period. This could be similar to the epidemic in some subgroups of MSM, such as young African American MSM.^{46, 47} When we solve our model for ART and PrEP in this type of epidemic, the optimal solution is again to spend the minimum amount on PrEP and invest the remaining budget in ART scale up. With our assumed budget and minimum acceptable levels of program implementation, there is excess budget after implementing up to x_1^{max} ($x_1=100,000$ incremental people receiving ART), so $x_2=28,414$ people receive PrEP. In this scenario, the ratio of PrEP health benefits, a_2 and b_2 , to ART benefits, a_1 and b_1 , is greater than in the base case. This is because prevention is relatively more important in a rapidly growing epidemic, where incidence is high and there are many potential infections to prevent. The same result is seen for CBE, but the greater number of infections averted per person in the rapidly growing epidemic leads to the program being cost-saving. In sensitivity analysis we did not consider the case of a diminishing epidemic, as there is less need and urgency for treatment and prevention investment.

As shown by these examples, the optimal balance of treatment and prevention depends heavily on the cost of the interventions. For PrEP, a very expensive preventive intervention, it is always optimal to invest only the minimum required amount in PrEP (if any) and spend

the remaining budget on scaling up ART. Even with very high PrEP efficacy, the high cost of PrEP precludes it from being a better value investment than ART. Similarly, CBE is a low-cost intervention, and it is optimal to invest in CBE before ART scale up unless CBE effectiveness is extremely low.

5. Discussion

Treatment scale up and prevention programs are both important in combating the HIV epidemic. Our model provides a simple yet accurate framework for determining the optimal mix of program investment.

Our analysis demonstrates that HIV budgets are often best spent on the program that offers the greatest “bang for the buck.” This is an intuitive finding, which is supplemented by the simple methods we outline for estimating the health benefits and costs of treatment and prevention programs. Decision makers could easily solve our model in a spreadsheet after collecting and estimating model inputs. The inputs, such as HIV incidence, prevalence, baseline ART use, and per person cost of prevention, are less extensive than for a dynamic model and draw from data to which public health decision makers would typically have access. Additionally, for the example analyses we carried out, the inclusion of program overlap did not change the optimal resource allocation, suggesting that good solutions can be obtained without estimating possible program overlap effects, as long as program coverage is not extremely high.

We illustrated our model with the example of ART, PrEP and CBE in MSM in the US. We saw that it is always better to invest in ART scale up rather than PrEP, because ART has a better ratio of health benefits to costs. Even with high PrEP efficacy or in a rapidly growing epidemic, factors that could tip the scales toward prevention, PrEP is a much less efficient use of resources than is ART scale up. This adds interesting context to the debate on PrEP use, especially given the ethical considerations of using ARVs for uninfected individuals when many infected individuals still need treatment.⁶¹ Community-based education is a much less costly intervention. With linear costs, it is best to invest as much as possible in such a program before investing in treatment scale up, but if the program has increasing marginal costs, a mix of investment in ART scale up and prevention may be optimal.

We consider the objectives of maximizing QALYs gained and HIV infections averted. For our example analyses, the choice of objective function did not change the optimal resource allocation. In other cases, particularly when the potential portfolio includes both prevention and treatment programs, the two objective functions may yield different solutions. Then QALYs gained may be the most appropriate objective to consider, as it comprehensively captures benefits of both treatment and prevention.³⁴

Our analysis has several limitations. Our framework accounts for net present costs of treatment and prevention and their future impact on the HIV epidemic. However, decision makers may think in terms of shorter time horizons and prefer to budget accordingly. Additionally, similar to most existing HIV resource allocation models, we have framed our problem as one of allocating a given budget among a predefined set of interventions, subject

to upper and lower limits on investment in each intervention. To determine the most efficient set of interventions to invest in, one must compare the costs and health benefits for each possible set of interventions. If desired, one can also use the framework to determine the diminution in health benefits (if any) arising from minimum required levels of investment in interventions. Such analysis is straightforward using our framework.

We approximate diminishing returns with nonlinear cost functions. Incorporating diminishing returns in the objective functions in addition to the cost functions could lead to a more accurate estimate of program effects. This would be a straightforward extension of the model but would affect analytical tractability. Additionally, we approximate dynamic effects with a static model. For the example problem of HIV control among MSM in the US, we showed that the resource allocation results from our static model (invest first in CBE, then ART, then PrEP) matched the results obtained using a more comprehensive dynamic model. For a very rapidly growing epidemic or when high levels of intervention coverage are considered, the static model may be a less accurate approximation of the epidemic dynamics and may overestimate health effects of interventions. The overestimation occurs because of diminishing epidemic returns as interventions are scaled up (the same infection cannot be prevented twice); such dynamic effects are not captured by the linear model. In such a case, use of a more complex dynamic model may be appropriate.

We designed our model to be applicable across different settings and for many types of HIV prevention programs. Although we developed it with the intention of comparing treatment and prevention programs, it could also be used to compare only prevention programs. Just as the basic structure facilitates comparing treatment to various types of prevention programs, one could estimate inputs for various prevention programs and solve for the optimal investment accordingly. Public health officials regularly must make decisions about how to allocate resources for HIV prevention, and our simple model could assist them in such decisions.

The simplicity of our model design also allows it to be used for different types of HIV epidemics (different growth rates or a different population). While populations targeted for HIV control are heterogeneous, with varied risk behaviors and mixing patterns, our model is intended to consolidate those factors into the model inputs, resulting in greater ease of use than with complex dynamic models. A decision maker who wishes to address different risk groups separately could utilize the model separately for each group (with appropriately estimated parameters), assuming the groups are independent.

HIV is a persistent burden in the US and worldwide, and only limited resources are available to control the epidemic. Public health decision makers need simple tools that help determine the optimal balance of investment in treatment and prevention. Our model can provide a simple yet accurate basis for determining optimal investment.

Acknowledgment

Financial support for this study was provided entirely by Grant Number R01-DA15612 from the National Institute on Drug Abuse. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

References

1. UNAIDS Report on the Global AIDS Epidemic 2013. Geneva, Switzerland: UNAIDS; 2013. Joint United Nations Programme on HIV/AIDS (UNAIDS).
2. Antiretroviral Therapy Cohort Collaboration. Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet*. 2008 Jul 26; 372(9635):293–299. [PubMed: 18657708]
3. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med*. 2011 Aug 11; 365(6):493–505. [PubMed: 21767103]
4. May M, Sterne JA, Sabin C, Costagliola D, Justice AC, Thiebaut R, et al. Prognosis of HIV-1-infected patients up to 5 years after initiation of HAART: collaborative analysis of prospective studies. *AIDS*. 2007 May 31; 21(9):1185–1197. [PubMed: 17502729]
5. Gebo KA, Fleishman JA, Conviser R, Hellinger J, Hellinger FJ, Josephs JS, et al. Contemporary costs of HIV healthcare in the HAART era. *AIDS*. 2010 Nov 13; 24(17):2705–2715. [PubMed: 20859193]
6. World Health Organization (WHO). Global HIV/AIDS Response: Epidemic Update and Health Sector Progress Towards Universal Access: Progress Report 2011. Geneva (Switzerland): World Health Organization; 2011.
7. Teshale, EH.; Kamimoto, L.; Harris, N.; Li, J.; Wang, H.; McKenna, MT. Estimated number of HIV-infected persons eligible for and receiving HIV antiretroviral therapy, 2003--United States (Abstract #167); 12th Conference on Retroviruses and Opportunistic Infections; 2005.
8. Bedimo AL, Pinkerton SD, Cohen DA, Gray B, Farley TA. Condom distribution: a costutility analysis. *Int J STD AIDS*. 2002 Jun; 13(6):384–392. [PubMed: 12015012]
9. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N Engl J Med*. 2010 Dec 30; 363(27):2587–2599. [PubMed: 21091279]
10. Juusola JL, Brandeau ML, Owens DK, Bendavid E. The cost-effectiveness of preexposure prophylaxis for HIV prevention in the United States in men who have sex with men. *Ann Intern Med*. 2012 Apr 17; 156(8):541–550. [PubMed: 22508731]
11. Cambiano V, Phillips AN. Modelling the impact of treatment with individual antiretrovirals. *Curr Opin HIV AIDS*. 2011 Mar; 6(2):124–130. [PubMed: 21505387]
12. Cohen DA, Wu SY, Farley TA. Comparing the cost-effectiveness of HIV prevention interventions. *J Acquir Immune Defic Syndr*. 2004 Nov 1; 37(3):1404–1414. [PubMed: 15483470]
13. Johnson LF, White PJ. A review of mathematical models of HIV/AIDS interventions and their implications for policy. *Sex Transm Infect*. 2011 Dec; 87(7):629–634. [PubMed: 21685191]
14. Alistar SS, Brandeau ML. Decision making for HIV prevention and treatment scale up: bridging the gap between theory and practice. *Med Decis Making*. 2012 Jan-Feb; 32(1):105–117. [PubMed: 21191118]
15. Brandeau ML, Zaric GS. Optimal investment in HIV prevention programs: more is not always better. *Health Care Manag Sci*. 2009 Mar; 12(1):27–37. [PubMed: 19938440]
16. Earnshaw SR, Hicks K, Richter A, Honeycutt A. A linear programming model for allocating HIV prevention funds with state agencies: a pilot study. *Health Care Manag Sci*. 2007 Sep; 10(3):239–252. [PubMed: 17695135]
17. Lasry A, Sansom SL, Hicks KA, Uzunangelov V. A model for allocating CDC's HIV prevention resources in the United States. *Health Care Manag Sci*. 2011 Mar; 14(1):115–124. [PubMed: 21184183]
18. Ruiz, M.; Gable, A.; Kaplan, EH.; Stoto, M.; Fineberg, H.; Trussell, J., editors. No Time to Lose: Getting More from HIV Prevention. Washington, DC: National Academy Press; 2001.
19. Gray RH, Li X, Wawer MJ, Gange SJ, Serwadda D, Sewankambo NK, et al. Stochastic simulation of the impact of antiretroviral therapy and HIV vaccines on HIV transmission; Rakai, Uganda. *AIDS*. 2003 Sep 5; 17(13):1941–1951. [PubMed: 12960827]

20. Lou J, Wu J, Chen L, Ruan Y, Shao Y. A sex-role-preference model for HIV transmission among men who have sex with men in China. *BMC Public Health*. 2009; 9(Suppl 1):S10. [PubMed: 19922680]
21. Nyabadza F. A mathematical model for combating HIV/AIDS in Southern Africa: Will multiple strategies work? *J Biol Syst*. 2006; 14(3):357–372.
22. Cremin I, Alsallaq R, Dybul M, Piot P, Garnett G, Hallett TB. The new role of antiretrovirals in combination HIV prevention: a mathematical modelling analysis. *AIDS*. 2013 Jan 28; 27(3):447–458. [PubMed: 23296196]
23. Hogan DR, Baltussen R, Hayashi C, Lauer JA, Salomon JA. Cost effectiveness analysis of strategies to combat HIV/AIDS in developing countries. *BMJ*. 2005 Dec 17; 331(7530):1431–1437. [PubMed: 16282380]
24. Bärnighausen T, Bloom DE, Humair S. Economics of antiretroviral treatment vs. circumcision for HIV prevention. *Proc Natl Acad Sci U S A*. 2012 Dec 26; 109(52):21271–21276. [PubMed: 23223563]
25. Long EF, Stavert RR. Portfolios of biomedical HIV interventions in South Africa: a cost-effectiveness analysis. *J Gen Intern Med*. 2013 Oct; 28(10):1294–1301. [PubMed: 23588668]
26. Alistar SS, Owens DK, Brandeau ML. Effectiveness and cost effectiveness of oral preexposure prophylaxis in a portfolio of prevention programs for injection drug users in mixed HIV epidemics. *PLoS One*. 2014 Jan 28; 9(1):e86584. [PubMed: 24489747]
27. Paltiel AD, Freedberg KA, Scott CA, Schackman BR, Losina E, Wang B, et al. HIV preexposure prophylaxis in the United States: impact on lifetime infection risk, clinical outcomes, and cost-effectiveness. *Clin Infect Dis*. 2009 Mar 15; 48(6):806–815. [PubMed: 19193111]
28. Paltiel AD, Weinstein MC, Kimmel AD, Seage GR 3rd, Losina E, Zhang H, et al. Expanded screening for HIV in the United States—an analysis of cost-effectiveness. *N Engl J Med*. 2005 Feb 10; 352(6):586–595. [PubMed: 15703423]
29. Walensky RP, Weinstein MC, Kimmel AD, Seage GR 3rd, Losina E, Sax PE, et al. Routine human immunodeficiency virus testing: an economic evaluation of current guidelines. *Am J Med*. 2005 Mar; 118(3):292–300. [PubMed: 15745728]
30. Burgos JL, Gaebler JA, Strathdee SA, Lozada R, Staines H, Patterson TL. Cost-effectiveness of an intervention to reduce HIV/STI incidence and promote condom use among female sex workers in the Mexico-US border region. *PLoS One*. 2010; 5(6):e11413. [PubMed: 20617193]
31. Hallett TB, Baeten JM, Heffron R, Barnabas R, de Bruyn G, Cremin I, et al. Optimal uses of antiretrovirals for prevention in HIV-1 serodiscordant heterosexual couples in South Africa: a modelling study. *PLoS Med*. 2011 Nov; 8(11):e1001123. [PubMed: 22110407]
32. Sanders GD, Bayoumi AM, Sundaram V, Bilir SP, Neukermans CP, Rydzak CE, et al. Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. *N Engl J Med*. 2005 Feb 10; 352(6):570–585. [PubMed: 15703422]
33. Ruiz, MS.; Gable, AR.; Kaplan, EH.; Stoto, MA.; Fineberg, HV.; Trussell, J., editors. *No Time to Lose: Getting More from HIV Prevention*. Washington, DC: Institute of Medicine; 2001.
34. Gold, MR.; Siegel, JE.; Russell, LB.; Weinstein, MC., editors. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
35. Kerr CC, Stuart RM, Gray RT, Shattock A, Fraser N, Benedikt C, et al. Optima: a model for HIV epidemic analysis, program prioritization, and resource optimization. *J Acquir Immune Defic Syndr*. 2015 Mar 23. Epub ahead of print.
36. Hillier, FS.; Lieberman, GJ. *Introduction to Operations Research*. 9th ed.. New York: McGraw-Hill; 2009.
37. Alistar SS, Owens DK, Brandeau ML. Effectiveness and cost effectiveness of expanding harm reduction and antiretroviral therapy in a mixed HIV epidemic: a modeling analysis for Ukraine. *PLoS Med*. 2011 Mar; 8(3):e1000423. [PubMed: 21390264]
38. Luenberger, DG.; Ye, Y. *Linear and Nonlinear Programming*. 3rd ed.. New York: Springer; 2010.
39. Marseille E, Dandona L, Marshall N, Gaist P, Bautista-Arredondo S, Rollins B, et al. HIV prevention costs and program scale: data from the PANCEA project in five low and middle-income countries. *BMC Health Serv Res*. 2007 Jul 12; 7:108. [PubMed: 17626616]

40. Kaplan EH. Economic analysis of needle exchange. *AIDS*. 1995; 9(10):1113–1120. [PubMed: 8519446]
41. Hutton DW, Brandeau ML. Too much of a good thing? When to stop catch-up vaccination. *Med Decis Making*. 2013 Oct; 33(7):920–936. [PubMed: 23858015]
42. Johns B, Torres TT. Costs of scaling up health interventions: a systematic review. *Health Policy Plan*. 2005 Jan; 20(1):1–13. [PubMed: 15689425]
43. Alistar, SS.; Brandeau, ML.; Beck, EJ. REACH (Resource Allocation for Control of HIV): a practical HIV resource allocation tool for decision makers. In: Zaric, GS., editor. *Operations Research and Health Care Policy*. New York: Springer; 2013. p. 201–224.
44. Dandona L, Kumar SP, Ramesh Y, Rao MC, Kumar AA, Marseille E, et al. Changing cost of HIV interventions in the context of scaling-up in India. *AIDS*. 2008; 22(Suppl 1):S43–S49. [PubMed: 18664952]
45. Dandona L, Sisodia P, Kumar SP, Ramesh Y, Kumar AA, Rao MC, et al. HIV prevention programmes for female sex workers in Andhra Pradesh, India: outputs, cost and efficiency. *BMC Public Health*. 2005; 5(1):98. [PubMed: 16181491]
46. Centers for Disease Control and Prevention (CDC). [cited 2012 8 March] Estimates of new HIV infections in the United States. 2008. Available from: <http://www.cdc.gov/nchhstp/newsroom/docs/Fact-Sheet-on-HIV-Estimates.pdf>
47. Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F, et al. Estimated HIV incidence in the United States, 2006–2009. *PLoS One*. 2011; 6(8):e17502. [PubMed: 21826193]
48. Desai K, Sansom SL, Ackers ML, Stewart SR, Hall HI, Hu DJ, et al. Modeling the impact of HIV chemoprophylaxis strategies among men who have sex with men in the United States: HIV infections prevented and cost-effectiveness. *AIDS*. 2008 Sep 12; 22(14):1829–1839. [PubMed: 18753932]
49. Kegeles SM, Hays RB, Pollack LM, Coates TJ. Mobilizing young gay and bisexual men for HIV prevention: a two-community study. *AIDS*. 1999 Sep 10; 13(13):1753–1762. [PubMed: 10509578]
50. Kahn JG, Kegeles SM, Hays R, Beltzer N. Cost-effectiveness of the Mpowerment Project, a community-level intervention for young gay men. *J Acquir Immune Defic Syndr*. 2001 Aug 15; 27(5):482–491. [PubMed: 11511826]
51. Centers for Disease Control and Prevention (CDC). [cited 2011 5 April] HIV and AIDS among gay and bisexual men. 2010. Available from: <http://www.cdc.gov/nchhstp/newsroom/docs/FastFacts-MSM-FINAL508COMP.pdf>
52. U.S. Census Bureau Population Division. [cited 2011 5 April] National population projections, released 2008 (based on Census 2000). 2008. Available from: <http://www.census.gov/population/www/projections/downloadablefiles.html>
53. Centers for Disease Control and Prevention (CDC). [cited 2009 30 September] HIV prevalence estimates -- United States, 2006. 2008. Available from: <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5739a2.htm>
54. Xu F, Sternberg MR, Markowitz LE. Men who have sex with men in the United States: demographic and behavioral characteristics and prevalence of HIV and HSV-2 infection: results from National Health and Nutrition Examination Survey 2001–2006. *Sex Transm Dis*. 2010 Jun; 37(6):399–405. [PubMed: 20473245]
55. Kaiser Family Foundation (KFF). U.S. Federal Funding for HIV/AIDS: The President's FY 2012 Budget Request. 2011 Available from: <http://www.kff.org/hivaids/upload/7029-07.pdf>.
56. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N Engl J Med*. 2012 Aug 2; 367(5):399–410. [PubMed: 22784037]
57. Marrazzo, J.; Ramjee, G.; Nair, G., et al. 20th Conference on Retroviruses and Opportunistic Infections. Atlanta, GA: 2013 Mar 3–6. Pre-exposure prophylaxis for HIV in women: daily oral tenofovir, oral tenofovir/emtricitabine, or vaginal tenofovir gel in the VOICE study (MTN 003) [Abstract]. 2013.

58. Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE, Segolodi TM, et al. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. *N Engl J Med*. 2012; 367(5):423–434. [PubMed: 22784038]
59. Van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J, Kapiga S, et al. Preexposure prophylaxis for HIV infection among African women. *N Engl J Med*. 2012 Aug 2; 367(5):411–422. [PubMed: 22784040]
60. Choopanya K, Martin M, Suntharasamai P, Sangkum U, Mock PA, Leethochawalit M, et al. Antiretroviral prophylaxis for HIV infection among people who inject drugs in Bangkok, Thailand: a randomized, double-blind, placebo-controlled trial. *Lancet*. 2013 Jun 15; 381(9883):2083–2090. [PubMed: 23769234]
61. Leibowitz AA, Parker KB, Rotheram-Borus MJ. A US policy perspective on oral preexposure prophylaxis for HIV. *Am J Public Health*. 2011 Jun; 101(6):982–985. [PubMed: 21493945]
62. Long EF, Owens DK. The cost-effectiveness of a modestly effective HIV vaccine in the United States. *Vaccine*. 2011 Aug 18; 29(36):6113–6124. [PubMed: 21510996]
63. McCormick AW, Walensky RP, Lipsitch M, Losina E, Hsu H, Weinstein MC, et al. The effect of antiretroviral therapy on secondary transmission of HIV among men who have sex with men. *Clin Infect Dis*. 2007 Apr 15; 44(8):1115–1122. [PubMed: 17366461]
64. Wilson DP, Law MG, Grulich AE, Cooper DA, Kaldor JM. Relation between HIV viral load and infectiousness: a model-based analysis. *Lancet*. 2008 Jul 26; 372(9635):314–320. [PubMed: 18657710]
65. Juusola JL, Brandeau ML, Long EF, Owens DK, Bendavid E. The cost effectiveness of symptom-based testing and routine screening for acute HIV infection in men who have sex with men in the United States. *AIDS*. 2011 Sep 10; 25(14):1779–1787. [PubMed: 21716076]

APPENDIX

Optimality Conditions for the Case of an ART Program, One Prevention Program, a Cobb-Douglas Overlap Function, and Increasing Marginal Costs

Consider the case of ART and one prevention program, with cost functions given by

$$C_i(x_i) = \begin{cases} C_i x_i, & x_i^{\min} \leq x_i \leq \frac{\beta_i}{\alpha_i} \\ C_i x_i + (\alpha_i x_i - \beta_i) x_i, & \frac{\beta_i}{\alpha_i} \leq x_i \leq x_i^{\max} \end{cases}$$

and overlap functions given by $\delta_Q(\mathbf{x}) = k_Q x_1 x_2$ and $\delta_{IA}(\mathbf{x}) = k_{IA} x_1 x_2$, where k_Q and k_{IA} are nonnegative constants. The resource allocation problem P3 is then:

$$\max_{x_1, x_2} Q(x_1, x_2) = a_1 x_1 + a_2 x_2 - k_Q x_1 x_2$$

or

$$\max_{x_1, x_2} IA(x_1, x_2) = b_1 x_1 + b_2 x_2 - k_{IA} x_1 x_2$$

$$s.t. C_1(x_1) + C_2(x_2) = B$$

$$x_1^{\min} \leq x_1 \leq x_1^{\max}$$

$$x_2^{\min} \leq x_2 \leq x_2^{\max}$$

Since the budget constraint is piecewise, we have four cases to consider:

$$(1)x_1 \leq \frac{\beta_1}{\alpha_1}, x_2 \leq \frac{\beta_2}{\alpha_2}; (2)x_1 \leq \frac{\beta_1}{\alpha_1}, x_2 > \frac{\beta_2}{\alpha_2}; (3)x_1 > \frac{\beta_1}{\alpha_1}, x_2 \leq \frac{\beta_2}{\alpha_2}; (4)x_1 > \frac{\beta_1}{\alpha_1}, x_2 > \frac{\beta_2}{\alpha_2}.$$

To solve this problem, we determine which cases are feasible based on the lower and upper limits for x_1 and x_2 . We then compare the optimal solutions for each feasible case.

In Case 1, the budget constraint is linear since x_1 and x_2 are small enough that diminishing returns do not yet take effect. Therefore Case 1 is simply P2. When we rearrange the budget constraint to obtain x_2 as a function of x_1 and substitute this into the objective functions,

$Q(x_1, x_2)$ and $IA(x_1, x_2)$ are strictly convex functions. We can show that $\frac{d^2 Q(x_1)}{dx_1^2} = 2k_Q \frac{C_1}{C_2}$ and $\frac{d^2 IA(x_1)}{dx_1^2} = 2k_{IA} \frac{C_1}{C_2}$, which are strictly greater than zero when overlap is present, since $k_Q > 0$ and $k_{IA} > 0$. Thus, the optimal solution is a boundary solution, as for P1. The same rules for optimally investing apply as in P1.

For Case 2, the budget constraint is $C_1 x_1 + C_2 x_2 + (\alpha_2 x_2 - \beta_2)x_2 = B$. We solve this for x_1 as a function of x_2 : $x_1 = \frac{B - C_2 x_2 - (\alpha_2 x_2 - \beta_2)x_2}{C_1}$. We then substitute to obtain an objective function of one variable. Our optimization problem becomes

$$\max_{x_2} Q(x_2) = a_1 \frac{B - C_2 x_2 - (\alpha_2 x_2 - \beta_2)x_2}{C_1} + a_2 x_2 - k_Q x_2 \frac{B - C_2 x_2 - (\alpha_2 x_2 - \beta_2)x_2}{C_1}$$

or

$$\max_{x_2} IA(x_2) = b_1 \frac{B - C_2 x_2 - (\alpha_2 x_2 - \beta_2)x_2}{C_1} + b_2 x_2 - k_{IA} x_2 \frac{B - C_2 x_2 - (\alpha_2 x_2 - \beta_2)x_2}{C_1}$$

$$s.t. \max \left(x_2^{\min}, \frac{\beta_2}{\alpha_2} \right) \leq x_2 \leq \min \left(x_2^{\max}, x_2^{\text{bound}} \right)$$

where x_2^{bound} is the nonnegative minimum of $\frac{\beta_2 - C_2 \pm \sqrt{(C_2 - \beta_2)^2 - 4\alpha_2(C_1 x_1^{\min} - B)}}{2\alpha_2}$. The objective function is cubic, so its convexity depends on x_2 . For example, for

maximizing QALYs gained, the objective function is strictly convex when

$x_2 > \frac{1}{3} \left(\frac{a_1}{k_Q} - \frac{C_2 - b_2}{\alpha_2} \right)$. Because of the nature of the objective function, rather than determining whether the function is convex or concave in the feasible range of x_2 for Case 2, for simplicity we check the optimality of all feasible boundary and interior solutions to determine which solution is optimal. We evaluate the objective function at the following solutions (x_1^*, x_2^*) :

Solution type	x_1^*	x_2^*
Lower boundary	$\min \left(x_1^{max}, \frac{B - C_2 x_2^* - (\alpha_2 x_2^* - \beta_2) x_2^*}{C_1} \right)$	$\max \left(x_2^{min}, \frac{\beta_2}{\alpha_2} \right)$
Upper boundary	$\min \left(x_1^{max}, \frac{B - C_2 x_2^* - (\alpha_2 x_2^* - \beta_2) x_2^*}{C_1} \right)$	$\min(x_2^{max}, x_2^{bound})$
Interior	$\min \left(x_1^{max}, \frac{B - C_2 x_2^* - (\alpha_2 x_2^* - \beta_2) x_2^*}{C_1} \right)$	$\min(x_2^{int})$
Interior	$\min \left(x_1^{max}, \frac{B - C_2 x_2^* - (\alpha_2 x_2^* - \beta_2) x_2^*}{C_1} \right)$	$\max(x_2^{int})$

where

$$x_2^{int} = \frac{-(k_Q C_2 - k_Q \beta_2 - a_1 \alpha_2) \pm \sqrt{(k_Q C_2 - k_Q \beta_2)^2 + k_Q \alpha_2 [3(k_Q B - a_2 C_1) + a_1 (C_2 - \beta_2)] + a_1^2 \alpha_1^2}}{3k_Q \alpha_2}$$

or

$$x_2^{int} = \frac{-(k_{IA} C_2 - k_{IA} \beta_2 - b_1 \alpha_2) \pm \sqrt{(k_{IA} C_2 - k_{IA} \beta_2)^2 + k_{IA} \alpha_2 [3(k_{IA} B - b_2 C_1) + b_1 (C_2 - \beta_2)] + b_1^2 \alpha_1^2}}{3k_{IA} \alpha_2}$$

and the feasibility of both values of x_2^{int} would need to be confirmed.

Case 3 is very similar to Case 2, but the budget constraint is quadratic in x_1 rather than x_2 : $C_1 x_1 + (\alpha_1 x_1 - \beta_1) x_1 + C_2 x_2 = B$. We follow the same steps in solving the problem as in Case 2. We first solve the budget constraint for x_2 as a function of x_1 :

$x_2 = \frac{B - C_1 x_1 - (\alpha_1 x_1 - \beta_1) x_1}{C_2}$. We substitute to obtain an objective function of one variable, and our optimization problem becomes

$$\max_{x_1} Q(x_1) = a_2 \frac{B - C_1 x_1 - (\alpha_1 x_1 - \beta_1) x_1}{C_2} + a_1 x_1 - k_Q x_1 \frac{B - C_1 x_1 - (\alpha_1 x_1 - \beta_1) x_1}{C_2}$$

or

$$\max_{x_1} IA(x_1) = b_2 \frac{B - C_1 x_1 - (\alpha_1 x_1 - \beta_1) x_1}{C_2} + b_1 x_1 - k_{IA} x_1 \frac{B - C_1 x_1 - (\alpha_1 x_1 - \beta_1) x_1}{C_2}$$

$$s.t. \max \left(x_1^{\min}, \frac{\beta_1}{\alpha_1} \right) \leq x_1 \leq \min \left(x_1^{\max}, x_1^{\text{bound}} \right)$$

where x_1^{bound} is the nonnegative minimum of $\frac{\beta_1 - C_1 \pm \sqrt{(C_1 - \beta_1)^2 - 4\alpha_1(C_2 x_2^{\min} - B)}}{2\alpha_1}$.

The convexity of the objective function depends on x_1 because it is cubic. For example, for

maximizing QALYs gained it is strictly convex when $x_1 > \frac{1}{3} \left(\frac{a_2}{k_Q} - \frac{C_1 - b_1}{\alpha_1} \right)$. Just as with Case 2, we evaluate the objective function for the feasible boundary and interior solutions (x_1^*, x_2^*) to determine which solution is optimal. The solutions we check are as follows:

Solution type	x_1^*	x_2^*
Lower boundary	$\max \left(x_1^{\min}, \frac{\beta_1}{\alpha_1} \right)$	$\min \left(x_2^{\max}, \frac{B - C_1 x_1^* - (\alpha_1 x_1^* - \beta_1) x_1^*}{C_2} \right)$
Upper boundary	$\min(x_1^{\max}, x_1^{\text{bound}})$	$\min \left(x_2^{\max}, \frac{B - C_1 x_1^* - (\alpha_1 x_1^* - \beta_1) x_1^*}{C_2} \right)$
Interior	$\min(x_1^{\text{int}})$	$\min \left(x_2^{\max}, \frac{B - C_1 x_1^* - (\alpha_1 x_1^* - \beta_1) x_1^*}{C_2} \right)$
Interior	$\max(x_1^{\text{int}})$	$\min \left(x_2^{\max}, \frac{B - C_1 x_1^* - (\alpha_1 x_1^* - \beta_1) x_1^*}{C_2} \right)$

where the term x_1^{bound} is the nonnegative minimum of

$$\frac{\beta_1 - C_1 \pm \sqrt{(C_1 - \beta_1)^2 - 4\alpha_1(C_2 x_2^{\min} - B)}}{2\alpha_1} \text{ and}$$

$$x_1^{\text{int}} = \frac{-(k_Q C_1 - k_Q \beta_1 - a_2 \alpha_1) \pm \sqrt{(k_Q C_1 - k_Q \beta_1 - a_2 \alpha_1)^2 - 3k_Q \alpha_1 (a_1 C_2 - a_2 (C_1 - \beta_1) - k_Q B)}}{3k_Q \alpha_1}$$

or

$$x_1^{int} = \frac{-(k_{IA}C_1 - k_{IA}\beta_1 - b_2\alpha_1) \pm \sqrt{(k_{IA}C_1 - k_{IA}\beta_1 - b_2\alpha_1)^2 - 3k_{IA}\alpha_1(b_1C_2 - b_2(C_1 - \beta_1) - k_{IA}B)}}{3k_{IA}\alpha_1}$$

and the feasibility of both values of x_1^{int} must be confirmed.

For Case 4, the budget constraint is quadratic in both in x_1 and x_2 : $C_1x_1 + (\alpha_1x_1 - \beta_1)x_1 + C_2x_2 + (\alpha_2x_2 - \beta_2)x_2 = B$. This constraint is strictly convex:

$$\frac{d^2C}{dx_1^2} = 2\alpha_1 > 0$$

$$\frac{d^2C}{dx_2^2} = 2\alpha_2 > 0$$

$$\frac{d^2C}{dx_1^2} \frac{d^2C}{dx_2^2} - \left[\frac{d^2C}{dx_1 dx_2} \right]^2 = 4\alpha_1\alpha_2 > 0$$

The objective function, $Q(x_1, x_2) = a_1x_1 + a_2x_2 - k_Qx_1x_2$ or $IA(x_1, x_2) = b_1x_1 + b_2x_2 - k_{IA}x_1x_2$, is neither convex nor concave:

$$\frac{d^2Q}{dx_1^2} = 0$$

$$\frac{d^2Q}{dx_2^2} = 0$$

$$\frac{d^2Q}{dx_1^2} \frac{d^2Q}{dx_2^2} - \left[\frac{d^2Q}{dx_1 dx_2} \right]^2 = -(k_Q)^2 < 0$$

and

$$\frac{d^2IA}{dx_1^2} = 0$$

$$\frac{d^2 IA}{dx_2^2} = 0$$

$$\frac{d^2 IA}{dx_1^2} \frac{d^2 IA}{dx_2^2} - \left[\frac{d^2 IA}{dx_1 dx_2} \right]^2 = -(k_{IA})^2 < 0$$

Thus this is a general case of constrained optimization, and we must use the first-order conditions to solve it. We determine the first-order conditions and solve for interior solutions, and then we test the optimality of feasible boundary and interior solutions (x_1^*, x_2^*) . The solutions we check are as follows:

Solution type	x_1^*	x_2^*
Lower boundary, x_1	$\max\left(x_1^{\min}, \frac{\beta_1}{\alpha_1}\right)$	$\min(x_2^{\max}, x_2^{\text{bound}})$
Lower boundary, x_2	$\min(x_1^{\max}, x_1^{\text{bound}})$	$\max\left(x_2^{\min}, \frac{\beta_2}{\alpha_2}\right)$
Interior	Numerically solve for (x_1^*, x_2^*) : $-2k_Q \alpha_1 (x_1^*)^2 + (2a_2 \alpha_1 + k_Q (\beta_1 - C_1)) x_1^* + a_2 (C_1 - \beta_1)$ $= -2k_Q \alpha_2 (x_2^*)^2 + (2a_1 \alpha_2 + k_Q (\beta_2 - C_2)) x_2^* + a_1 (C_2 - \beta_2)$ or $-2k_{IA} \alpha_1 (x_1^*)^2 + (2b_2 \alpha_1 + k_{IA} (\beta_1 - C_1)) x_1^* + b_2 (C_1 - \beta_1)$ $= -2k_{IA} \alpha_2 (x_2^*)^2 + (2b_1 \alpha_2 + k_{IA} (\beta_2 - C_2)) x_2^* + b_1 (C_2 - \beta_2)$ $x_2^* = \frac{\beta_2 - C_2 \pm \sqrt{(C_2 - \beta_2)^2 - 4\alpha_2 [\alpha_1 (x_1^*)^2 + (C_1 - \beta_1) x_1^* - B]}}{2\alpha_2}$	

where the term x_1^{bound} is the nonnegative minimum of

$$\frac{\beta_1 - C_1 \pm \sqrt{(C_1 - \beta_1)^2 - 4\alpha_1 [\alpha_2 (x_2^{\min})^2 + (C_2 - \beta_2) x_2^{\min} - B]}}{2\alpha_1} \text{ and } x_2^{\text{bound}} \text{ is the}$$

nonnegative minimum of

$$\frac{\beta_2 - C_2 \pm \sqrt{(C_2 - \beta_2)^2 - 4\alpha_2 [\alpha_1 (x_1^{\min})^2 + (C_1 - \beta_1) x_1^{\min} - B]}}{2\alpha_2}.$$

Estimation of Overlap Factors for the Example of ART, PrEP, and CBE in the US

We used simulation results from a dynamic model of treatment and prevention among MSM in the US¹⁰ to develop estimation methods for k_Q and k_{IA} . We determined that the constants k_Q and k_{IA} depend on HIV incidence in the population, efficacy of treatment and prevention in terms of a_1 and b_1 and ε_i , time horizon T , and the initial size of the susceptible population,

for $k_Q = \frac{(\text{average incidence}) \times a_1 \times \varepsilon_i \times (33.5) \ln(T)}{(\# \text{initially susceptible})}$ and $k_{IA} = \frac{(\text{average incidence}) \times b_1 \times \varepsilon_i \times (3t - (7570)(\text{average incidence}) + 188)}{(\# \text{initially susceptible})}$. Overlap is inversely proportional to the initial size of the susceptible population. This is intuitive: when the susceptible population is large, treatment and prevention programs will be spread among many people, whereas when the susceptible population is small, both programs are trying to protect the same people. Overlap increases with a_1 and b_1 and ε_i , which is also intuitive because the magnitude of the overlap depends on the magnitude of health benefits from each intervention. QALYs gained are slightly less affected by overlap than are infections averted, since some fraction of the QALYs gained are due to benefits to the individuals being treated, and those QALYs are not affected by concurrent implementation of a prevention program.

When solving the two-program case of P2, we assume the budget would always be fully utilized. One can imagine that if k_Q or k_{IA} were large enough, the objective function would begin to decrease at some threshold of x_1 and x_2 and so health benefits would be greater if implementation did not go beyond that point. However, because k_Q and k_{IA} are likely to be very small relative to a_i and b_i , we assume that increasing x_1 and x_2 always yields positive health outcomes, and thus the budget constraint would always be binding.

Figure 1.

Linear estimates of health benefits from investments in ART and PrEP for MSM in the US. Estimated health benefits for each level of implementation of x_1 , the incremental number of people receiving ART, are plotted for P2. Any budget remaining after x_1 is implemented is invested in x_2 , PrEP. Only feasible values of x_1 are considered (e.g., from $x_1^{\min}=50,000$ to the level of x_1 achievable at $x_2^{\min}$). Figure 1a shows QALYs gained, Figure 1b shows infections averted. ART = antiretroviral therapy; PrEP = preexposure prophylaxis; QALY = quality-adjusted life year.

Figure 2.

Estimated health benefits from investment in antiretroviral therapy (ART, Figure 2a), preexposure prophylaxis (PrEP, Figure 2b), and community-based education (CBE, Figure 2c) for MSM in the US calculated by a linear model (P1; solid lines) and a dynamic model (dashed lines). Estimated health benefits are plotted for separate ART, PrEP, and CBE programs as determined by the linear model in this study (P1) and a previously published dynamic model. Figure 2a shows QALYs gained and infections averted for ART, Figure 2b

shows QALYs gained and infections averted for PrEP, and Figure 2c shows QALYs gained and infections averted for CBE. ART = antiretroviral therapy; CBE = community-based education; PrEP = preexposure prophylaxis; QALY = quality-adjusted life year.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3.

Estimated net present costs and QALYs over a 20-year time horizon from investment in all possible combinations of ART, PrEP, and CBE for MSM in the US calculated by a linear model (P1). Upper and lower limits on investment in each intervention are as described in the text. ART = antiretroviral therapy; CBE = community-based education; PrEP = preexposure prophylaxis; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

Figure 4. Estimate of QALYs gained (Figure 4a) and HIV infections averted (Figure 4b) for ART and PrEP for MSM in the US as a function of time horizon of program implementation. Estimated per person health benefits (a_1 , a_2 , b_1 , b_2) are plotted for ART and PrEP as a

function of the time horizon of program implementation. The percentage of the benefits from 20-year implementation is noted for $T=5$. Figure 4a shows QALYs gained (a_1 , a_2), Figure 4b shows infections averted (b_1 , b_2). ART = antiretroviral therapy; PrEP = preexposure prophylaxis; QALY = quality-adjusted life year.

Table 1

Model notation

Variable	Description
T	Time horizon for program implementation, years
x_1	Number of incremental people starting ART over T years due to scale up; $0 \leq x_1^{min} \leq x_1 \leq x_1^{max}$
x_i	Number of uninfected people reached by prevention program i ($i = 2, \dots, n$) over T years; $0 \leq x_i^{min} \leq x_i \leq x_i^{max}$
$\mathbf{x}$	Number of people starting ART and reached by prevention programs; $(x_1, x_2, \dots, x_{n-1}, x_n)$
$Q(\mathbf{x})$	Net present number of QALYs gained from the implementation of $\mathbf{x}$
$IA(\mathbf{x})$	Net present number of HIV infections averted from the implementation of $\mathbf{x}$
a_1	Net present number of QALYs gained per person starting ART; $a_1 > 0$
a_i	Net present number of QALYs gained per person reached by prevention program i ; $a_i > 0$
b_1	Net present number of infections averted per person starting ART; $b_1 > 0$
b_i	Net present number of infections averted per person reached by prevention program i ; $b_i > 0$
q_{ix}	Net present number of QALYs a treated individual gains from receiving ART
q_{IA}	Net present number of QALYs an individual gains from avoiding infection
σ_T	Chance that a person acquires HIV in T years
ε_{ix}	Efficacy of ART in reducing sexual infectivity in infected individuals
ε_i	Efficacy of prevention intervention i in reducing the chance of HIV acquisition
$C(\mathbf{x})$	Net present healthcare costs in the modeled population with the implementation of $\mathbf{x}$
B	Budget for HIV spending in the modeled population
C_1	Net present cost of starting one additional person on ART
C_i	Net present cost of reaching one uninfected person with prevention program i for T years
s_{IA}	Net present savings from averting an HIV infection
c_1	Net present lifetime costs of ART for one person
c_i	Net present cost of prevention program i for one person for T years
$\delta_Q(\mathbf{x})$	Magnitude of overlap in QALYs gained by the implementation of $\mathbf{x}$
$\delta_{IA}(\mathbf{x})$	Magnitude of overlap in infections averted by the implementation of $\mathbf{x}$
k_Q	Overlap constant for QALYs gained; $k_Q > 0$
k_{IA}	Overlap constant for infections averted; $k_{IA} > 0$
$\lambda_1(x_1)$	Average cost above baseline per person receiving ART at program coverage of x_1 ; assumed to be nonnegative, nondecreasing, and linear
$\lambda_i(x_i)$	Average cost above baseline per person in prevention program i at coverage of x_i ; assumed to be nonnegative, nondecreasing, and linear
α_i	Slope of the $\lambda_i(x_i)$ function
β_i	Intercept of the $\lambda_i(x_i)$ function
r	Annual discount rate for costs and health benefits
D	Cumulative discount factor for a T -year time horizon

Note: ART = antiretroviral therapy; QALY = quality-adjusted life year

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Parameter estimation guidelines *

Health Benefit Parameters	
$D = \sum_{t=1}^T \frac{1}{(1+r)^t}$ Discount factor: Treatment: $a_1 \approx q_{tx} + q_{IA}(b_1/D)$ Prevention: $a_i \approx q_{IA}(\sigma_T \varepsilon_i + (b_i/D)), i = 2, \dots, n$	
Estimation Guidelines	
q_{tx}	Net present number of QALYs a treated individual gains from receiving ART; estimates available in literature or can be estimated from existing models
q_{IA}	Net present number of QALYs an individual gains from avoiding infection; estimates available in literature or can be estimated from existing models
σ_T	Chance that a person acquires HIV in T years Stable epidemic: $\sigma_T \approx 1 - \left(\frac{\text{\#uninfected at end of year}}{\text{\#uninfected at start of year}} \right)^T$ Growing epidemic: $\sigma_T \approx \frac{1 - \left(\frac{\text{\#uninfected at end of year } j}{\text{\#uninfected at start of year } j} \right)^T + 1 - \left(\frac{\text{\#uninfected at end of year } k}{\text{\#uninfected at start of year } k} \right)^T}{2}$
ε_{tx}	Efficacy of ART in reducing sexual infectivity in infected individuals; estimates available in literature
ε_i	Efficacy of prevention intervention i in reducing the chance of HIV acquisition; estimates available in literature
b_1	Net present number of infections averted per person starting ART Stable epidemic: $b_1 \approx \left(\frac{\text{\#new infections per year}}{\text{\#untreated} + (1 - \varepsilon_{tx}) \times \text{\#treated}} - \frac{\text{\#new infections per year} \times (1 - \varepsilon_{tx})}{\text{\#untreated} + (1 - \varepsilon_{tx}) \times \text{\#treated}} \right) \times D$ $\approx \left[\left(\frac{\text{\#new infections in year } j}{\text{\#untreated in year } j + (1 - \varepsilon_{tx}) \times \text{\#treated in year } j} \right) + \frac{\text{\#new infections in year } k}{\text{\#untreated in year } k + (1 - \varepsilon_{tx}) \times \text{\#treated in year } k} \right] \times \frac{1}{2} D$ Growing epidemic: $b_1 \approx \left[\left(\frac{\text{\#new infections in year } j \times (1 - \varepsilon_{tx})}{\text{\#untreated in year } j + (1 - \varepsilon_{tx}) \times \text{\#treated in year } j} \right) + \frac{\text{\#new infections in year } k \times (1 - \varepsilon_{tx})}{\text{\#untreated in year } k + (1 - \varepsilon_{tx}) \times \text{\#treated in year } k} \right] \times \frac{1}{2} D$
b_i	Net present number of infections averted per person starting prevention program i Stable epidemic: $b_i \approx \sigma_T \varepsilon_i \left(\frac{\text{\#new infections per year}}{\text{\#infected at start of year}} \right) \times D$ Growing epidemic: $b_i \approx \sigma_T \varepsilon_i \left(\min \left(\left(\frac{\text{\#new infections in year } j}{\text{\#infected at start of year } j} \right), \left(\frac{\text{\#new infections in year } k}{\text{\#infected at start of year } k} \right) \right) \right) \times D$
Cost Parameters	
$D = \sum_{t=1}^T \frac{1}{(1+r)^t}$ Discount factor: Treatment: $C_1 \approx c_1 - s_{IA}(b_1/D)$ Prevention: $C_i \approx c_i - s_{IA}(\sigma_T \varepsilon_i + (b_i/D)), i = 2, \dots, n$	
Estimation Guidelines	
c_1	Difference in discounted lifetime healthcare costs for an infected individual receiving ART and an infected individual who never receives ART; estimates available in literature

c_i	Net present cost of prevention program i for one person for T years; estimates available in literature or can be calculated based on program components
s_{IA}	Difference in discounted lifetime healthcare costs for an infected individual and an uninfected individual; estimates available in literature

Note: ART = antiretroviral therapy; QALY = quality-adjusted life year

* A number of simplifying assumptions are made in creating the parameter estimates. QALY gains from treatment (a_1) are estimated as the direct QALY gains to the individual being treated plus estimated secondary QALY gains accruing from other individuals in the population who thereby avoid infection. The net present QALY gains from prevention (a_i) are estimated as the direct QALY gains when the individual avoids infection plus estimated secondary QALY gains when other individuals thereby avoid infection. The chance a person acquires HIV in T years (σ_T) is a cumulative probability calculated without conditioning on whether the individual survives previous periods. The net present number of infections averted per person starting ART (b_1) is calculated as the difference in the number of infections caused per year by untreated individuals versus treated individuals, multiplied by the length of the time horizon and the discount factor. The net present number of infections averted per person starting prevention program i (b_i) is calculated assuming an average rate of new infections per year, multiplied by the annual chance a person acquires HIV infection, the efficacy of the prevention program, the length of the time horizon, and the discount factor. The net present per person cost of treatment (C_1) is calculated as the cost of treatment for the person who enrolls in treatment minus an estimate of averted treatment costs from individuals who thereby avoid infection. The net present per person cost of a prevention program (C_i) is calculated as the cost of the program minus an estimate of averted treatment costs from an individual who avoids infection and other individuals who thereby avoid infection.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Key parameter values for numerical examples of antiretroviral therapy (ART) and prevention for MSM (comprising either preexposure prophylaxis (PrEP) or community-based education (CBE))

Parameter	PrEP and ART	CBE and ART	Source
Time Horizon			
Time horizon for program implementation, years, T	20	20	Assumed
Program Constraints			
Minimum acceptable ART scale up level, x_1^{min}	50,000	50,000	Assumed
Maximum feasible ART scale up level, x_1^{max}	400,000	400,000	Assumed
Minimum acceptable prevention implementation level, x_2^{min}	10,000	10,000	Assumed
Maximum feasible prevention implementation level, x_2^{max}	3,000,000	3,000,000	Assumed
Health Benefits			
Net present QALYs gained per person starting ART, a_1	15.53	15.53	Calculated (see Section 3.1)
Net present QALYs gained per person reached by prevention program, a_2	1.24	0.222	Calculated (see Section 3.1)
Net present infections averted per person starting ART, b_1	0.873	0.873	Calculated ^{47, 51, 53} (see Section 3.1)
Net present infections averted per person reached by prevention program, b_2	0.054	0.010	Calculated ^{9, 47, 49, 51, 53} (see Section 3.1)
Net present QALYs gained from receiving ART, q_{ix}	5	5	10
Net present QALYs gained from avoiding infection, q_{IA}	9	9	10, 62
Chance of acquiring HIV in T years, σ_T	14.6%	14.6%	Calculated ^{47, 51, 52} (see Section 3.1)
ART efficacy in reducing sexual infectivity, ε_{ix}	90%	90%	3, 63–65
Prevention efficacy in reducing the chance of HIV acquisition, ε_2	44%	8%	9, 49
Costs			
Budget for HIV spending, B	\$10 billion	\$10 billion	Estimated ⁵⁵
Net present cost of starting one person on ART, C_1	\$118,100	\$118,100	Calculated (see Section 3.2)
Net present cost of reaching one person with prevention program for T years, C_2	\$140,393	\$272	Calculated (see Section 3.2)
Net present savings from averting an HIV infection, s_{IA}	\$70,000	\$70,000	10, 62
Net present lifetime cost of ART for one person, c_1	\$200,000	\$200,000	10
Net present cost of prevention program for one person for T years, c_2	\$150,000	\$2,000	Calculated ^{10, 50} (see Section 3.2)
Overlap constant* for QALYs gained, k_Q	1.44E-06	2.62E-07	Calculated (see Appendix)
Overlap constant* for infections averted, k_{IA}	1.51E-07	2.75E-08	Calculated (see Appendix)

Parameter	PrEP and ART	CBE and ART	Source
Cost of ART above baseline, $\lambda_1(x_1)$	0	0	Assumed
Cost of prevention above baseline, $\lambda_2(x_2)$	.05 x_2 –300 for $x_2 > 6,000$	.01 x_2 –200 for $x_2 > 20,000$	Estimated ^{39, 42}
Discount rate for costs, QALYs and new infections, r	3%	3%	34

Note: ART = antiretroviral therapy; PrEP = preexposure prophylaxis; CBE = community-based education; QALY = quality-adjusted life year

* Assumes overlap functions of the form $\delta_Q(\mathbf{x}) = k_Q x_1 x_2$ and $\delta_{IA}(\mathbf{x}) = k_{IA} x_1 x_2$.

Table 4

Key parameter values for sensitivity analysis of numerical example of antiretroviral therapy (ART) and preexposure prophylaxis (PrEP) for MSM

Parameter	High PrEP Efficacy	Shorter Time Horizon	Rapidly Growing Epidemic	Source
Time Horizon				
Time horizon for program implementation, years, T	20	5	20	Assumed
Program Constraints				
Minimum acceptable ART scale up level, x_1^{min}	50,000	10,000	10,000	Assumed
Maximum feasible ART scale up level, x_1^{max}	400,000	100,000	100,000	Assumed
Minimum acceptable PrEP implementation level, x_2^{min}	10,000	2,000	1,000	Assumed
Maximum feasible PrEP implementation level, x_2^{max}	3,000,000	1,000,000	500,000	Assumed
Health Benefits				
Net present QALYs gained per person starting ART, a_1	15.53	7.64	28.94	Calculated
Net present QALYs gained per person starting PrEP, a_2	2.04	0.199	2.91	Calculated
Net present infections averted per person starting ART, b_1	0.873	0.269	1.98	Calculated ^{47, 51, 53}
Net present infections averted per person starting PrEP, b_2	0.089	0.004	0.160	Calculated ^{9, 47, 49, 51, 53}
Net present QALYs gained from receiving ART, q_{rx}	5	5	5	10
Net present QALYs gained from avoiding infection, q_{IA}	9	9	9	10, 62
Chance of acquiring HIV in T years, σ_T	14.6%	3.9%	24.5%	Calculated ^{47, 51, 52}
ART efficacy in reducing sexual infectivity, ε_{rx}	90%	90%	90%	3, 63, 64
PrEP efficacy in reducing the chance of HIV acquisition, ε_2	73%	44%	44%	9
Costs				
Budget for HIV spending, B	\$10 billion	\$2.5 billion	\$5 billion	Estimated ⁵⁵
Net present cost of starting one person on ART, C_1	\$118,100	\$179,490	\$13,800	Calculated
Net present cost of starting one person on PrEP for T years, C_2	\$134,139	\$44,449	\$127,404	Calculated
Net present savings from averting an HIV infection, s_{IA}	\$70,000	\$70,000	\$70,000	10, 62
Net present lifetime cost of ART for one person, c_1	\$200,000	\$200,000	\$200,000	10
Net present cost of PrEP for one person for T years, c_2	\$150,000	\$46,000	\$150,000	Calculated ¹⁰
Overlap constant* for QALYs gained, k_Q	2.39E-06	3.80E-07	2.32E-05	Calculated (see Appendix)
Overlap constant* for infections averted, k_{IA}	2.51E-07	3.54E-08	2.13E-06	Calculated (see Appendix)
Discount rate for costs, QALYs and new infections, r	3%	3%	3%	34

Note: ART = antiretroviral therapy; PrEP = preexposure prophylaxis; QALY = quality-adjusted life year

* Assumes overlap functions of the form $\delta_Q(\mathbf{x}) = k_Q x_1 x_2$ and $\delta_{IA}(\mathbf{x}) = k_{IA} x_1 x_2$.