

Optima: a model for HIV epidemic analysis, program prioritization, and resource optimization

Cliff C. Kerr^{1,2}, Robyn M. Stuart^{1,3}, Richard T. Gray¹, Andrew Shattock¹, Nicole Fraser⁴, Clemens Benedikt⁴, Markus Haacker^{4,5}, Maxim Berdnikov⁶, Ahmed Mohamed Mahmood⁷, Seham Abdalla Jaber⁷, Marelize Gorgens⁴, David P. Wilson¹

¹ Kirby Institute, University of New South Wales, Sydney, NSW, Australia

² School of Physics, University of Sydney, Sydney, NSW, Australia

³ Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

⁴ Global HIV/AIDS Program, The World Bank, Washington DC, USA

⁵ Harvard School of Public Health, Harvard University, Boston, MA, USA

⁶ The Global Fund, Geneva, Switzerland

⁷ Sudan National HIV/AIDS Control Programme, Khartoum, Sudan

Abstract

Optima is a software package for modeling HIV epidemics and interventions that we developed to address practical policy and program problems encountered by funders, governments, health planners, and program implementers. Optima's key feature is its ability to perform resource optimization to meet strategic HIV objectives, including HIV-related financial commitment projections and health economic assessments. Specifically, Optima allows users to choose a set of objectives (such as minimizing new infections, minimizing HIV-related deaths, and/or minimizing long-term financial commitments) and to then determine the optimal resource allocation (and thus program coverage levels) for meeting those objectives. These optimizations are based on: calibrations to epidemiological data; assumptions about the costs of program implementation and the corresponding coverage levels; and the effects of these programs on clinical, behavioral, and other epidemiological outcomes. Optima is flexible in terms of which population groups (specified by behavioral, epidemiological, and/or geographical factors) and which HIV programs are modeled, the amount of input data used, and the types of outputs generated. Here, we introduce this model and compare it to existing HIV models that have been used previously to inform decisions about HIV program funding and coverage targets. Optima has already been used in over 20 countries, and there is increasing demand from stakeholders to have a tool that can perform evidence-based HIV epidemic analyses, revise and prioritize national strategies based on available resources, set program coverage targets, amend subnational program implementation plans, and inform the investment strategies of governments and their funding partners.

Introduction

Despite decades of research and investment, HIV remains a major contributor to the global burden of disease [1]: each day, over 7,000 people become newly infected with HIV [2]. AIDS remains one of the leading causes of deaths for adults globally and particularly in sub-Saharan Africa [3]. Furthermore, development assistance for HIV has stabilized and further increases of HIV financing are likely to be from domestic sources [4,5]. In this context of limited resources, it

is imperative to allocate available funds as efficiently as possible. Especially as countries transition to domestically funded programs, they can consider different implementation approaches to reduce unit costs.

An allocatively efficient HIV response is one in which the funding for HIV programs (and thus the coverage levels for each program) is allocated in a way that will yield the greatest impact in managing and reducing HIV disease burden over a specified time period. Quantifiable improvements in HIV allocative efficiency have long been sought [6]. While most countries acknowledge the need for allocative efficiency, investing in the right mix of programs for the right populations in the right geographical areas is challenging, due to both political interests and the complexity of most HIV epidemics. Consequently, many countries do not prioritize the most efficacious interventions or scale them to appropriate coverage levels [7]. Recently, Anderson et al. [8] performed a detailed allocative efficiency analysis of the Kenyan HIV epidemic that examined resource allocations across multiple populations, interventions, and geographical locations. They found that 14% more infections could be averted over the study period (2014–2029) if resources were targeted to the most effective interventions and the regions most in need.

To help national governments and other stakeholders understand their HIV epidemics and allocate limited resources most efficiently, we developed Optima (formerly known as Prevtool; e.g., [9,10]), a software toolbox that models (a) HIV transmission within and between population groups, (b) disease progression, (c) the effects of HIV prevention and treatment programs, and (d) the economic effects of policy choices. We designed it to be flexible and comprehensive enough to accommodate the regional, national, and epidemiological diversity of HIV epidemics. Optima can be used to (a) estimate epidemiological trends to produce long-term forecasts, including for counterfactual scenarios; (b) calculate program cost-effectiveness, returns on investment, and other economic and HIV-related health outcomes; (c) determine the allocation of resources and associated coverage levels that minimize any of several objectives, including the number of new infections, HIV-related deaths and disease burdens, current and future HIV-related costs, or combinations thereof; and (d) determine the minimal resources required to achieve specific targets regarding those objectives. Innovatively, it performs optimization of resource allocations over different time periods using a formal mathematical algorithm.

The Optima model includes sexual, injecting-related, and vertical transmission of HIV, and can incorporate an arbitrarily large number of different population groups, including key affected populations and age stratifications. In **Figure 1A**, we show a typical selection of population groups for a concentrated HIV epidemic. Here, seven populations are used, including males and females in the general population, sex workers and their clients, males and females who inject drugs, and men who have sex with men (MSM). Populations can also be stratified by age, which is particularly important for generalized epidemics. Optima allows flexible definitions of population groups: for example, all or some MSM may also engage in injecting behavior, or persons in the general population of different ages may have different sexual behavior and choose partners either younger or older than themselves. The number of HIV programs and their definitions are also not fixed: users can choose HIV programs for specific populations or even to try out the impacts of new HIV programs or new (lower-cost) HIV service delivery models. Optima can incorporate different HIV service delivery models, including different unit

cost estimates for given program coverage levels; potential impacts of technical (program) efficiency gains can thus be included in the analyses.

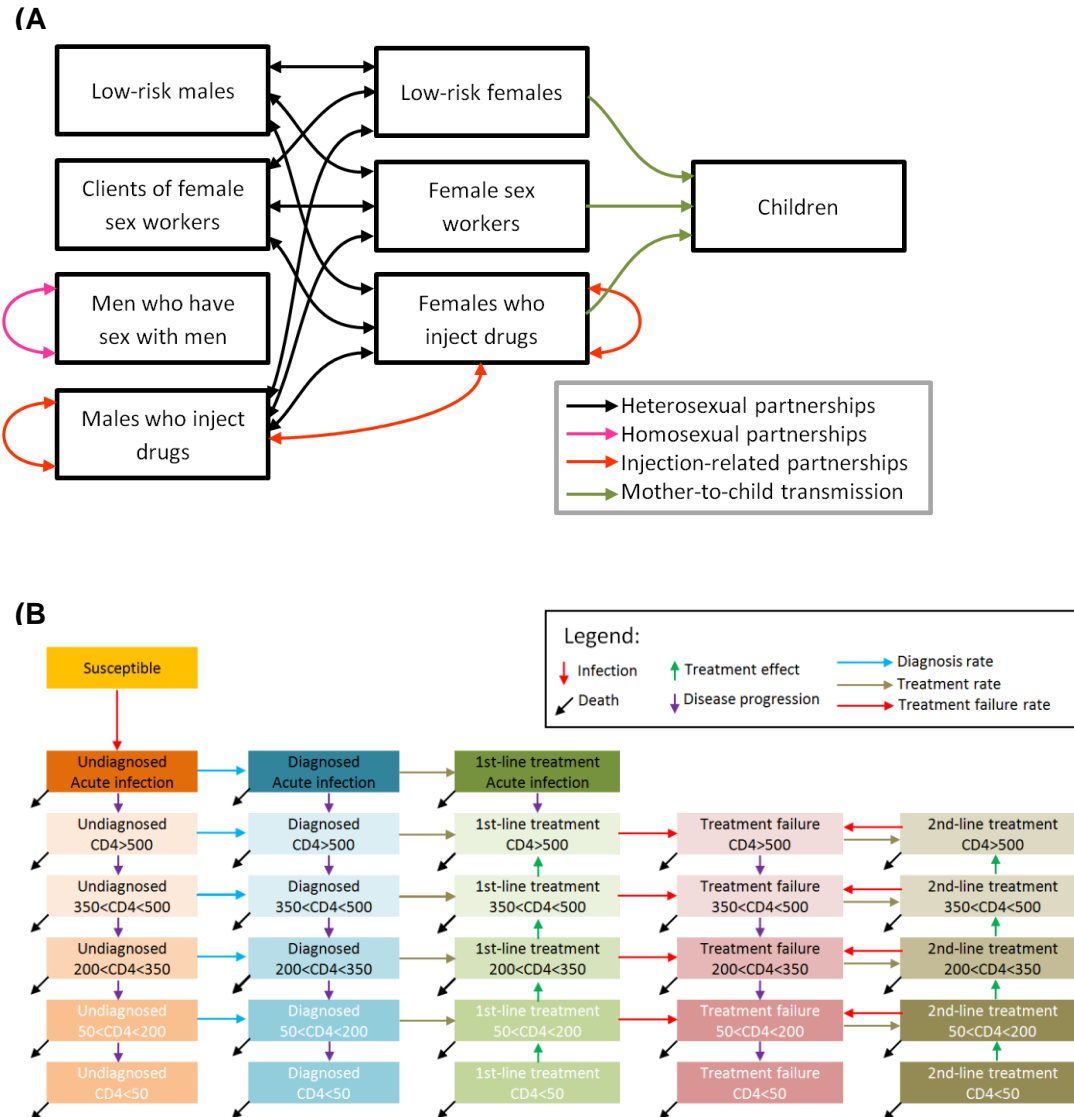


Figure 1: (A) Example population groups and HIV transmission-related interactions in Optima. (B) Schematic diagram of the health state structure of the model. Each compartment represents a single population group with the specified health state, while each arrow represents the movement of numbers of individuals between health states. All compartments except for “susceptible” represent individuals living with HIV. Death includes all causes of death.

We have so far completed allocative efficiency analyses with Optima in over 20 countries across Africa, Eastern Europe, Latin America and the Caribbean, and Asia. To our knowledge, Optima is the only HIV software package that allows users to optimize funding to meet strategic HIV program impacts without presupposing program coverage levels. This paper outlines the methodology underlying Optima and compares Optima to other commonly used HIV models, namely the Goals (Spectrum) Model[11,12], the AIDS Epidemic Model (AEM) [13], the Estimation and Projection Package (EPP) [14,15], and the Modes of Transmission (MOT) model

[16]. To illustrate Optima's use in the real world, we present a case study of how it was applied in Sudan, a low-income country with a low-level HIV epidemic.

Methods

This section provides a qualitative description of the methods used in Optima; further details are provided in the supplementary material.

HIV epidemic model

Optima is based on a dynamic, population-based HIV model; **Figure 1B** shows the disease progression implemented in the model. Optima tracks the entire population of people living with HIV (PLHIV) across five stages of CD4 count. These CD4 count stages are aligned to the progression of WHO treatment guidelines, namely acute HIV infection, >500 cells/ μ L, 350–500 cells/ μ L, 200–350 cells/ μ L, 50–200 cells/ μ L, and <50 cells/ μ L. Key aspects of the antiretroviral therapy (ART) service delivery cascade are included: from infection to diagnosis, ART initiation on first-line therapy, treatment failure, subsequent lines of therapy, and HIV/AIDS-related or other death. The primary purpose of HIV testing is to identify those who are HIV-positive. With the new UNAIDS global targets of 90% of PLHIV identified by 2020, 90% of them on treatment, and 90% of these virally suppressed [12], the structure of the disease progression model in Optima is designed to help countries measure and achieve this goal, and optimize resource allocations accordingly.

The model uses a linked system of ordinary differential equations to track the movement of PLHIV between HIV health states; the full set of equations is provided in the supplementary material. The overall population is partitioned in two ways: by population group and by HIV health state. Individuals are assigned to a given population group based on their dominant risk. However, to capture important cross-modal types of transmission, relevant behavioral parameters can be set to nonzero values (e.g., males who inject drugs may engage in commercial sex; some MSM may have female sexual partners).

HIV infections occur through the interaction between different populations via regular, casual, or commercial (including transactional) sexual partnerships, through sharing of injecting equipment, or through mother-to-child transmission. The force-of-infection is the rate at which uninfected individuals become infected, and it depends on the number and type of risk events to which individuals are exposed in a given period (either within their population groups or through interaction with other population groups) and the infection probability of each event. Mathematically, the force-of-infection has the general form:

$$\lambda = 1 - (1 - \beta)^n$$

where λ is the force-of-infection, β is the transmission probability of each event, and n is the effective number of at-risk events (i.e., n gives the average number of interaction events with HIV-infected people where HIV transmission may occur). The value of the transmission probability β varies across CD4 count compartments (indirectly reflecting the high viral load at early and late stages of infection), differs for different modes of transmission (intravenous drug injection with a contaminated needle-syringe, penile-vaginal or penile-anal intercourse, and mother-to-child), and may be reduced by behavioral interventions (for example, condom use)

[17,18], biological interventions (for example, male circumcision) or antiretroviral therapy [19,20]. There is one force-of-infection term for each type of interaction (for example, casual sexual relationships between males and female sex workers); the force-of-infection for a given population will be the sum of all interaction types.

For sexual transmission, the force-of-infection is determined by:

- The HIV prevalence (weighted by viral load) in partner populations;
- The average number of casual, regular, and commercial homosexual and heterosexual acts per person per year;
- The proportion of these acts in which condoms are used;
- The proportion of men who are circumcised;
- The prevalence of sexually transmissible infections (which can increase HIV transmission probability);
- The proportion of acts that are covered by pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis (PEP);
- The proportion of partners on antiretroviral treatment (ART); and
- The efficacies of condoms, male circumcision, PEP, PrEP, and ART at preventing HIV transmission.

For injecting-related transmission, the force-of-infection is determined by:

- The HIV prevalence (weighted by viral load) in populations of people who use a syringe and then share it;
- The number of injections per person per year;
- The proportion of injections that use shared equipment;
- The fraction of people who inject drugs on opioid substitution therapy and its efficacy in reducing injecting behavior.

For mother-to-child transmission (MTCT), the number of infections is determined by:

- The birth rate among women living with HIV;
- The proportion of women with HIV who breastfeed;
- The probability of perinatal HIV transmission in the absence of intervention; and
- The proportion of women receiving prevention of mother-to-child transmission (PMTCT), including ART.

In addition to the force-of-infection rate, which is the number of individuals who become infected with HIV per year, there are seven other ways individuals may change health states. First, individuals may die, either due to an average background death rate for that population (which is greater for older populations or for people who inject drugs) or due to HIV/AIDS (which depends on CD4 count). Second, in the absence of treatment, individuals progress from higher to lower CD4 counts. Third, individuals can move from undiagnosed to diagnosed states based on their HIV testing rate, which depends on CD4 count (for example, people with AIDS symptoms or primary HIV infection may have a higher testing rate) and population type (for example, female sex workers may test more frequently than males in the general population). Fourth, diagnosed individuals may commence ART, at a rate depending on CD4 count. Fifth, individuals may experience treatment failure due to lack of adherence to therapy or development of drug resistance, and sixth, people may initiate second and subsequent lines of

treatment from treatment failure. Finally, while on successful first- or second-line treatment (i.e. effective viral suppressive therapy), individuals may progress from lower to higher CD4 counts.

The change in the number of people in each compartment is determined by the sum over the relevant rates described above, multiplied by the population size of the compartments on which they act. For example, the change in the number of undiagnosed HIV-positive female sex workers (FSW) with a CD4 count between 200 and 350 cells/ μ L is:

$$\frac{dU_{FSW_{200-350}}}{dt} = U_{FSW_{350-500}}\tau_{350-500} - U_{FSW_{200-350}}(\mu_{200-350} + \tau_{200-350} + \eta_{FSW_{350-500}})$$

where $U_{FSW_{200-350}}$ is the current number of undiagnosed HIV-positive female sex workers with a CD4 count between 200 and 350 cells/ μ L, $U_{FSW_{350-500}}$ is the same population but with higher CD4 count (350–500 cells/ μ L), τ is the disease progression rate for the given CD4 count (where $1/\tau$ is the average time to lose 150 CD4 cells/ μ L), μ is the death rate, and η is the HIV testing rate. (Note: this example does not consider movement between populations, such as FSW returning to the general female population and vice versa – something which is included in Optima.) Each compartment (**Figure 1B**, boxes) corresponds to a single differential equation in the model, and each rate (**Figure 1B**, arrows) corresponds to a single term in that equation.

Table 1: Input parameters of the model.

	Biological parameters	Behavioral parameters	Epidemiological/other parameters
<i>Population parameters</i>	Background death rate		Population sizes (T, P)
<i>HIV-related parameters</i>	Sexual HIV transmissibilities* (H) STI-related transmissibility increase* Condom efficacy* Circumcision efficacy* HIV health state progression rates (H) HIV-related death rates (H)	Number of sexual partners* (T, P, S) Number of acts per partner* (S) Condom usage probability* (T, P) Circumcision probability* (T)	HIV prevalence (T, P) STI prevalence (T, P)
<i>MTCT parameters</i>	Mother-to-child transmission probability*	Birth rate* PMTCT access rate* (T)	
<i>Injection-related parameters</i>	Injecting HIV transmissibility* Syringe cleaning efficacy* Drug-related death rate	Number of injections* (T) Syringe sharing probability* (T) Syringe cleaning probability* Methadone treatment probability (T)	
<i>Treatment parameters</i>	ART efficacy in reducing infectiousness* ART failure rates	HIV testing rates (T, P, H)	Number of people on ART (T)
<i>Economic parameters</i>	Health utilities		Costs of all prevention, care and treatment programs, enablers and management (T, I) Cost-outcome curves (T, I) Discounting and inflation rates (T) Healthcare costs

Key: T = parameter value changes over time; P = parameter value depends on population group; H = parameter depends on health state; S = parameter depends on sexual partnership type; I = parameter depends on intervention type; * = parameter is used to calculate the force-of-infection.

Table 1 lists the parameters used in Optima; most of these are for calculating the force-of-infection. We interpret empirical estimates for model parameter values in Bayesian terms as prior distributions. The model must then be calibrated, which is the process of finding posterior distributions of the model parameter values such that the model generates accurate estimates of HIV prevalence, the number of people on treatment, and any other epidemiological data that

are available (e.g., HIV-related deaths). The calibration can be done automatically, manually, or a combination of both. This process of model calibration and validation should normally be done in consultation with governments in the countries in which the model is being applied.

HIV resource optimization and program coverage targets

A novel component of Optima is its ability to calculate allocations of resources that optimally address one or more HIV-related objectives (e.g., impact-level targets in a country's HIV national strategic plan). Because Optima also calculates the coverage levels required to achieve these targets, it can be used to inform HIV strategic planning and the determination of program coverage levels. The key assumptions of resource optimization are the relationships between (a) the cost of HIV programs for specific target populations, (b) the resulting coverage levels of targeted populations with these HIV programs, and (c) how these coverage levels of HIV programs for targeted populations influence behavioral and clinical outcomes. Such relationships are required to understand how incremental changes in spending (marginal costs) affect HIV epidemics. A traditional approach is to apply unit cost values to inform a linear relationship between money spent and coverage attained. This is a reasonable assumption for programs like an established ART program that no longer incurs start-up or initiation costs, but less appropriate for condom promotion and behavior change communication programs. Most HIV programs typically have initial setup costs, followed by a more effective scale-up with increased funding. However, there are saturation effects for very high coverage levels, since these require increased incremental costs due to demand generation and related activities for the most difficult-to-reach groups.

Optima uses a logistic function fitted to available input data to model cost-coverage curves; **Figure 2** shows an example. (Coverage-outcome relationships are assumed to be linear here for illustration purposes.) Logistic functions can incorporate initial startup costs and allow changes in behavior to saturate at high spending levels, thus better reflecting program reality. The logistic function has the form

$$L(x) = A + \frac{B-A}{1+e^{-(x-C)/D}},$$

where $L(x)$ relates spending to coverage, x is the amount of funding for the program, A is the lower asymptote value (adjusted to match the value of L when there is no spending on a program), B is the upper asymptote value (for very high spending), C is the midpoint, and D is the steepness of the transition from A to B . For our fits, we typically choose saturation values of the coverage to match behavioral data in countries with heavily funded HIV responses. Program coverage for zero spending, or behavioral outcomes for zero coverage of formal programs, are inferred using data from early on in the epidemic or just prior to significant investment in HIV programs. Practically, we also discuss the zero and high spending cases with local experts who can advise on private sector HIV service delivery outside the governments' expenditure tracking systems.

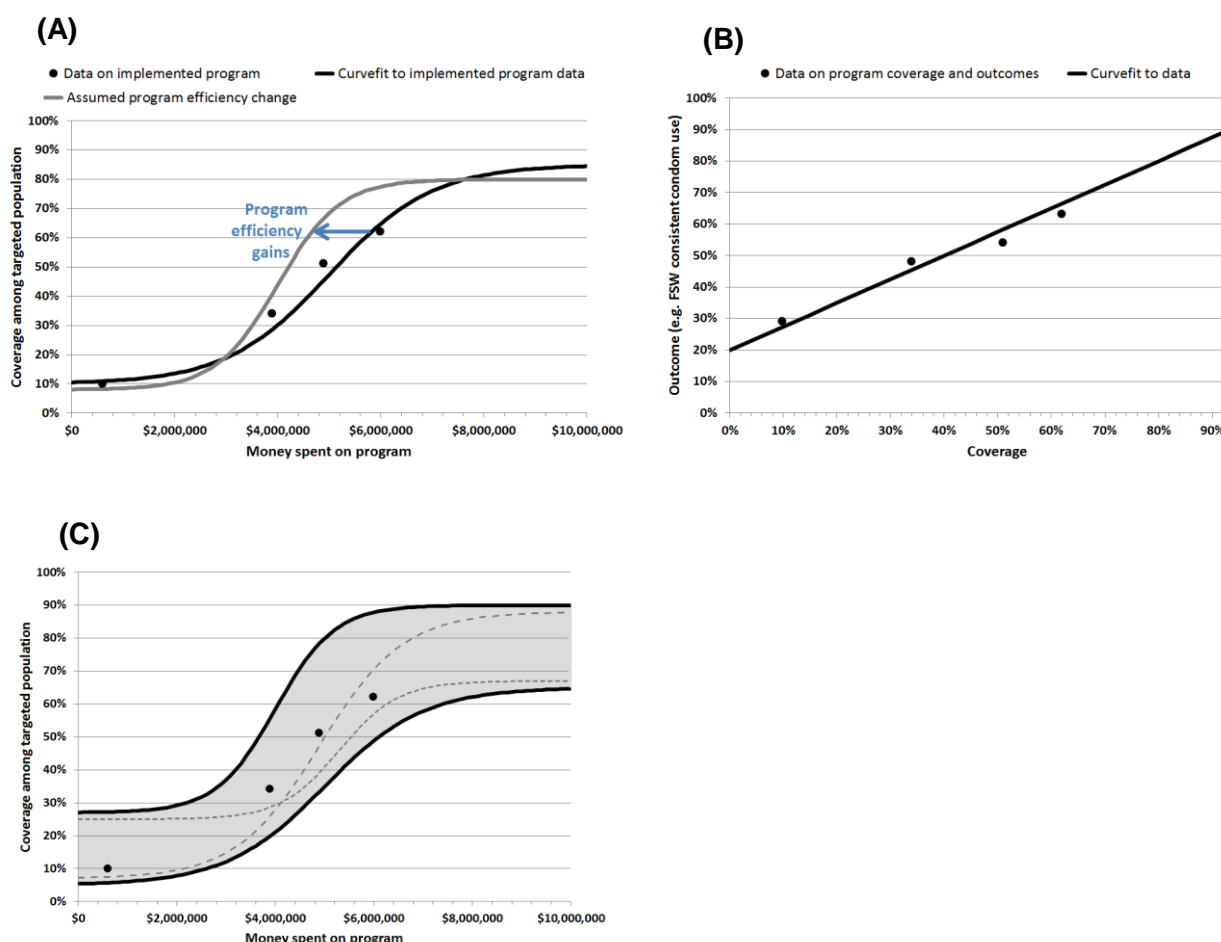


Figure 2: (A) Example cost-coverage curve, showing the relationship between spending on a program and the associated coverage attained among the target population, including technical efficiency gains. (B) Example coverage-outcome curve, showing the relationship between coverage of a minimum service package to a targeted population (e.g., outreach and condom distribution with minimum service package to female sex workers) and the resultant outcome (e.g., percentage of female sex workers consistently using condoms with their commercial partners). (C) Example of uncertainty incorporated in the cost-coverage curves used in Optima.

For each HIV program, we derive one set of logistic curves that relate funding to program coverage levels, and another set of curves (generally linear relationships) between coverage levels and clinical or behavioral outcomes (i.e., the impacts that HIV strategies aim to achieve). In future, Optima will include a default set of these cost-coverage-outcome curves, based on all available international evidence. Outcomes expected from changes in program funding are assumed by interpolating and extrapolating available data using a fitted logistic curve. A limitation of this approach is that all changes in behavior are assumed to be due to changes in program funding.

Optima can be used to minimize either (a) a given outcome (e.g., number of infections, number of disability-adjusted life years, number of HIV-related deaths, or future HIV-related costs) given a fixed total budget over a determined program period, or (b) the amount of funding required to meet a particular epidemiological goal (e.g., reducing HIV incidence by 50%). Optima can also determine the amount of money required to simultaneously meet multiple

goals (e.g. all impact-level targets in an HIV national strategic framework) or the optimal allocation of a fixed amount of resources which will simultaneously get as close as possible to achieving one or multiple target objectives. Optima can also be used to help decide in which geographic areas to implement programs for which target populations, or how to most effectively re-invest the savings from technical efficiency gains. Constraints may be placed on the optimization; for example, the number of people on antiretroviral therapy or prevention of mother-to-child transmission program coverage may not be allowed to decrease or programs cannot increase or decrease from a baseline level by more than a defined percentage each year to account for political or other constraints.

To perform the optimization, Optima uses a global parameter search algorithm called Bayesian adaptive locally linear stochastic descent (BALLSD) [21]. BALLSD is similar to simulated annealing in that it makes stochastic downhill steps in parameter space from an initial starting point. However, unlike simulated annealing, BALLSD chooses future step sizes and directions based on the outcome of previous steps. For certain classes of optimization problems, we have shown that BALLSD can determine optimal solutions with fewer function evaluations than traditional optimization methods, including gradient descent and simulated annealing [21].

Uncertainty analyses

Optima uses a Markov chain Monte Carlo (MCMC) algorithm [22-24] for performing automatic calibration and for computing uncertainties in the model fit to epidemiological data. With this algorithm, the model is run many (typically 1,000–10,000) times to generate a range of epidemic projections; their differences represent uncertainty in the expected epidemiological trajectories.

The most important assumptions in the optimization analysis are associated with the cost-coverage and coverage-outcome curves. To incorporate uncertainty in these curves, users define upper and lower limits for both coverage and behavior for no spending and for very high spending. All available historical spending data and achieved outcomes of spending, data from comparable settings, experience, and extensive discussion with stakeholders in the country of application can be used to inform these ranges. All logistic curves within these ranges are then allowable (see **Figure 2C**) and are incorporated into uncertainty analyses of Optima. These cost-coverage and coverage-outcome curves are thus reconciled with the epidemiological, behavioral, and biological data in a Bayesian-optimal way, thereby allowing the calculation of unified uncertainty estimates.

Time-varying optimization and time periods

Optima can also optimize program funding over time. The funding to each program is able to vary either (a) by a pre-determined year-by-year total budget, or (b) such that the total pool of funding is distributed across time in an optimal manner for the best overall outcome at the end of the analysis period. The optimal allocation of resources to each program may change over time in either situation. Time-varying program allocation is governed by the equation

$$A(t) = be^{st^2},$$

where A is the allocation of program funding as a function of time t , b determines the overall budget for a given program, and s is a shape parameter such that program funding is constant if

$s = 0$, increases over time if $s > 0$, and decreases if $s < 0$. The time points t are normalized such that the analysis period is mapped onto the closed interval $[0, 1]$.

Time periods are an important consideration in any resource allocation analysis. Some programs may be effective over the long-term but fail to receive funding if optimizations only consider the short-term, due to the time required to reach sufficient coverage and for effectiveness to become apparent (e.g., circumcision of male youth prior to sexual debut). Consequently, Optima allows two timeframes to be specified: the period over which the *funding* is to be optimized, and the period over which the *impacts* of said funding are to be optimized. For example, Optima can be used to calculate the optimal budget allocation for a four-year national strategic plan in terms of minimizing infections over the next 30 years. Of course, the two periods may align: for example, Optima can optimize funding over the next 10 years with respect to minimizing disability-adjusted life years (DALYs) over the same period [1]. This flexibility in time periods allows users to explore short- and long-term impacts of different investment strategies.

Technical efficiency gains

Technical (program) efficiency focuses on minimizing unit costs of overall service delivery, subject to community-level factors, the policy environment, and considerations regarding implementation quality. Equally important are administrative rules and regulations, which can hinder timely provision of resources from central to community levels. Understanding how management, financial analysis, and institutional efficiencies affect delivery costs can result in changes to service delivery models. These differences ultimately change the overall cost required to reach a target population with services of a given quality. Accordingly, program efficiency data and/or assumptions can be incorporated within Optima epidemic projections or allocative efficiency analyses by modifying the cost-coverage logistic curves; the initial start-up costs and/or slope of the curves can be adjusted to account for different program efficiency options (**Figure 2**). Optima can therefore show how to best allocate the savings incurred from less expensive service delivery to further maximize the achievement of HIV program impact targets.

Future financial commitments

Optima calculates the costs incurred per HIV infection, including costs for first and subsequent-lines of treatment, PMTCT, treatment of opportunistic infections, and care and treatment related to other HIV-related morbidities. It can also estimate the future financial liability to governments for the care and treatment of people currently living with HIV, as well as from people who are projected to become infected in the future.

From a public debt perspective, expressing this total cost as a percentage of annual GDP, over time, shows a government the total long-term financial implications of its HIV strategy and commitments. This is essential when (a) projecting long-term financial costs of HIV, (b) considering the financial sustainability and fiscal space available for HIV programs, and (c) determining whether resources may become available for non-HIV-related programs. HIV program financial sustainability is achieved when a country can reliably mobilize domestic and external resources to achieve the current and future coverage of HIV services necessary to achieve that country's HIV strategy goals. In the context of a rapidly changing AIDS funding landscape, it is important for governments to understand the long-term financial burdens of

their HIV epidemics, and consider how their HIV programs could be financed and sustained if donor funding were scaled down or became unavailable.

Results

Here we present a case study summarizing the findings of an allocative efficiency analysis of Sudan's national HIV epidemic and response. The analysis was conducted in support of the government's preparation of a concept note for submission to the Global Fund, the largest funding body in the nation's HIV response. Two key policy questions were formulated in consultation with the Sudanese National AIDS Program:

1. Finding the optimal allocation of HIV funds to minimize either cumulative HIV incidence by 2020 or cumulative HIV-related disability-adjusted life years (DALYs) by 2020; and
2. Determining the HIV funding required to achieve impact targets that were either moderate (25% reduction in HIV incidence and deaths by 2020 compared to 2010 levels) or ambitious (50% reduction in HIV incidence and deaths by 2020 compared to 2010 levels).

Data to inform the modeling were gathered from a comprehensive literature review and validated in consultation with key in-country and donor stakeholders. Ten populations were defined for inclusion in the model: female sex workers (FSW); their clients (SWC); men who have sex with men (MSM); children; and youth (aged 15-24 years), adults (aged 25-49 years), and older people (aged 50+ years), each disaggregated by sex. Seven core HIV programs were identified for optimization: programs for FSW; programs for MSM; programs for SWC; general population condom programs; HIV testing and counseling for the general population; ART; and prevention of mother-to-child transmission (PMTCT). Expenditure from the National AIDS Spending Assessment (NASA) for 2013 was used as a baseline; total spending was 12.2 million USD, with spending on these seven programs of 6.4 million USD.

The results addressing the first key policy question established that by optimally reallocating the same 6.4 million USD in programmatic spending as in 2013, Sudan could avert an additional 19,000 HIV infections (or 36% of cumulative new HIV infections) between 2014 and 2020. The simulations showed that to minimize incidence or DALYs, scaling up the ART program is highest priority, followed by targeted primary prevention programs for FSW, their clients, and MSM, with some adjustments to the allocation given to each of these programs depending on the choice of strategic objective (Figure 3). Should more programmatic funding become available – for example, by mobilizing additional funds or reducing management costs – HIV incidence (or DALYs) could be reduced further (Figure 4).

In answer to the second key policy question, the modeling results showed that the minimum annual programmatic spending required to achieve the moderate impact targets outlined in the National Strategic Plan was 8.1 million USD. Achieving the more ambitious targets of a 50% reduction in HIV incidence and AIDS-related deaths by 2020 would require an estimated 34 million USD annually for programs – over five times the spend of USD 6.4 million in 2013. The cost of reaching HIV incidence reduction targets is estimated to be lower than cost of reaching HIV mortality targets, which require extensive coverage of HIV testing programs in the general population.

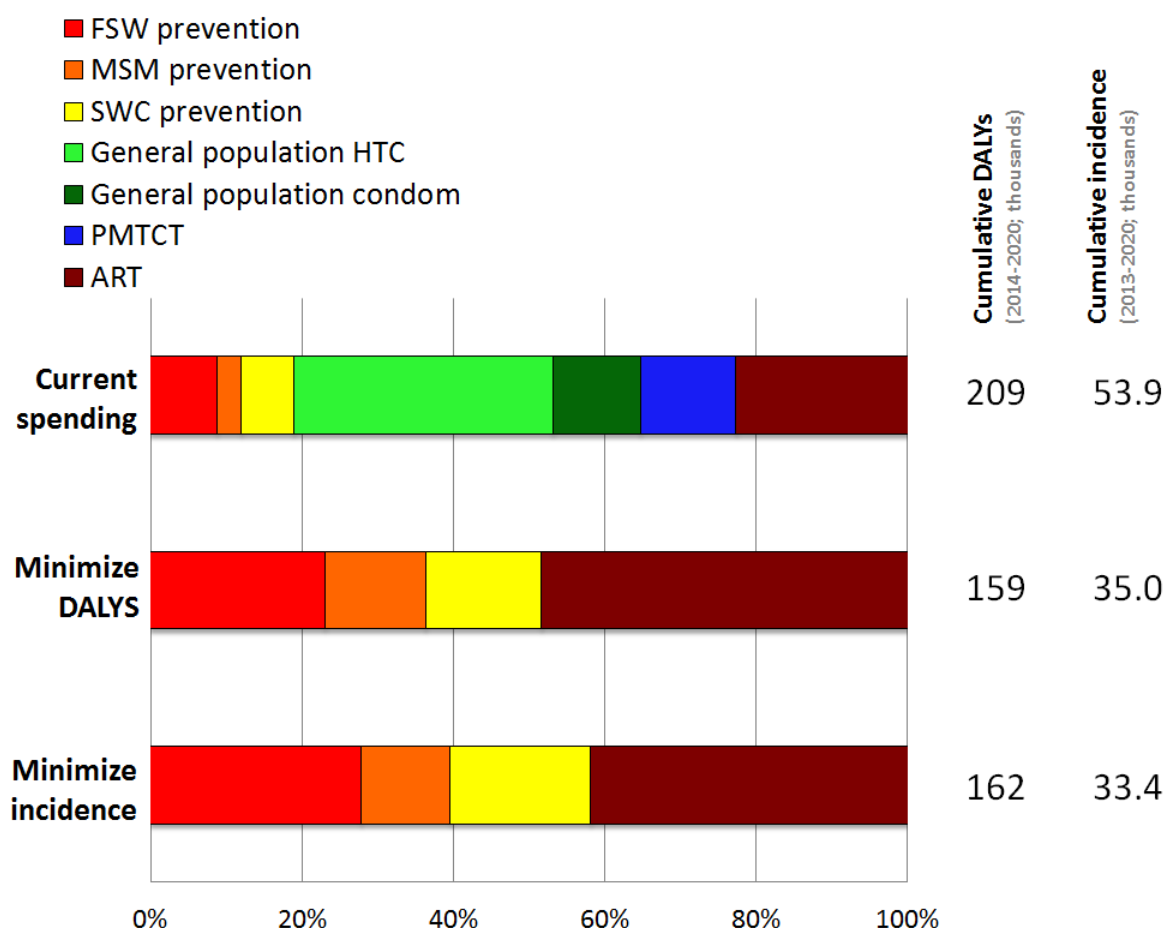


Figure 3: Current (2013) versus optimal allocations of HIV resources in Sudan for two different optimization objectives.

A key finding of all the analyses conducted was that targeting programmatic resources almost exclusively to ART and prevention programs for FSW, SWC, and MSM has downstream effects on all population groups and leads to HIV incidence reductions in all populations over the medium term, and at lower cost than targeting the general population directly. This is as expected in a low-level HIV epidemic, where transmission in the general population is largely fueled by new infections occurring in the key populations.

The allocative efficiency study carried out in Sudan broadly highlighted some of Optima's ability to provide policy-relevant information for high-level decision making and planning. Based on these results, the Sudanese government shifted their program priorities and budgets in their request for funding from the Global Fund.

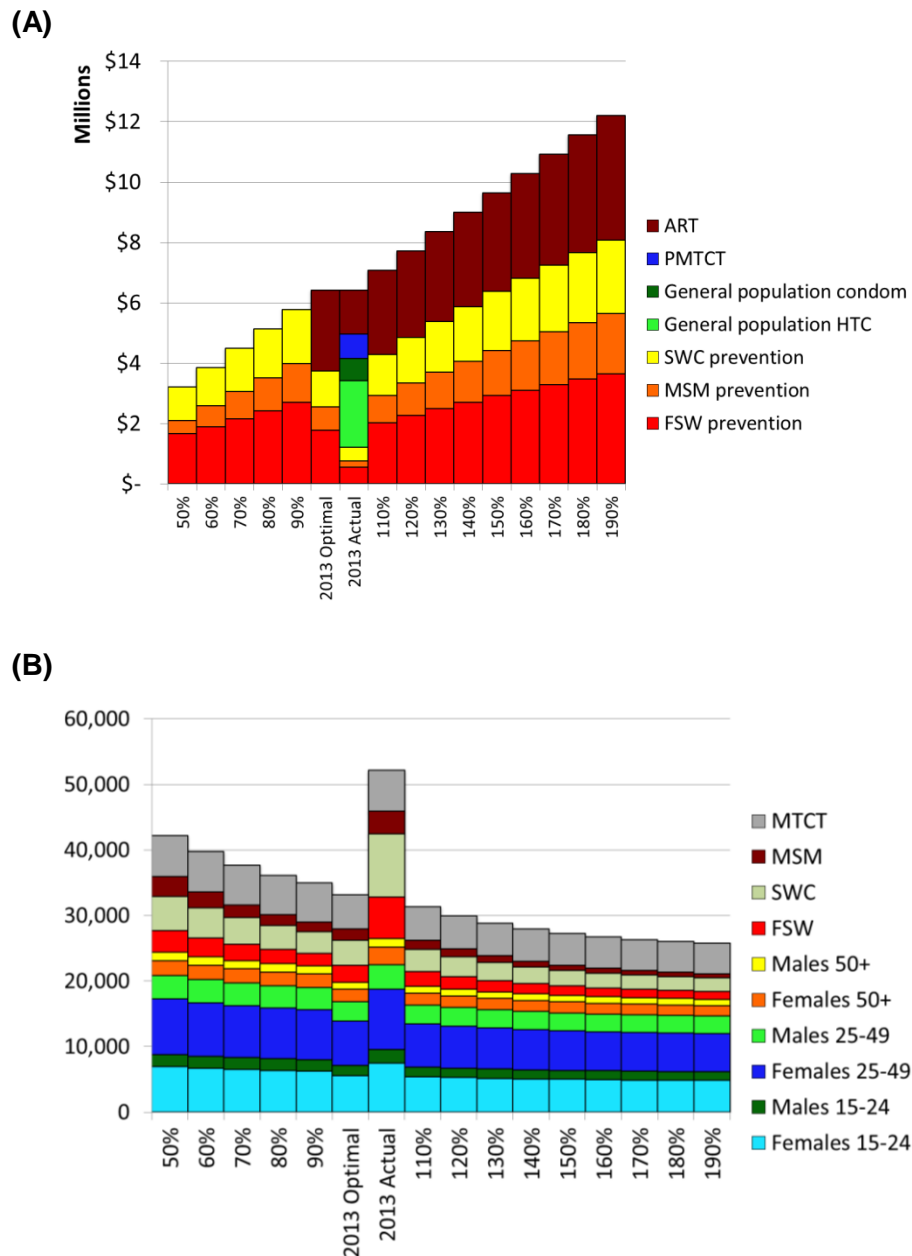


Figure 4: (A) Resource allocations optimized to minimize HIV incidence by 2020 at different budget levels for Sudan. Prevention packages for key populations include outreach, condoms, and HTC but do not include ART (which is shown separately). (B) The model-estimated number of cumulative infections resulting from the allocations shown in Figure 4A.

Discussion

Optima has been designed to answer the key questions that arise as national governments and other HIV stakeholders decide how to fund their HIV responses and make choices regarding allocations and target program coverage levels. Specifically, it allows them to:

1. Understand what population groups and behaviors are chiefly responsible for new HIV infections in the relevant settings, and how these change over time;

2. Project current trends and understand what the likely trajectory of their HIV epidemics are if current conditions continue;
3. Explore the epidemiological, health, and economic benefits that past investments in packages of HIV programs have likely had, which may be used as either an evaluation or an advocacy tool;
4. Predict the outcomes of different possible future funding scenarios, such as scenarios that prioritize treatment over prevention or vice versa, or certain subpopulations or geographical regions over others;
5. Formally and quantitatively determine the funding allocation given the currently available budget that minimizes specific epidemiological or economic indicators (e.g., new HIV infections, DALYs, HIV-related deaths, or long-term HIV-related health care costs);
6. Estimate the minimum resources and coverage targets required to meet specific epidemiological, economic, and program impact targets, including multiple targets (e.g., 90% reductions in both incidence and DALYs by 2030);
7. Estimate the impacts and marginal effects of new programs being introduced (provided that basic information about the program's efficacy at the individual level is known);
8. Estimate where to best allocate savings from technical efficiency gains;
9. Understand how optimal responses may change over time by front-loading some programs and then reducing them while scaling up others;
10. Explore geospatial optimized allocations and priority program coverage;
11. Conduct analyses over different time periods to explore the short- and long-term impacts of certain investment strategies over a specific period (such as over a national strategy); and
12. Calculate the long-term financial commitment caused by one new HIV infection, as well as annual and total future costs for care and treatment of people living with HIV.

A number of alternative software packages for modeling HIV epidemics and HIV program impacts and costs already exist. Several of the models that have historically been used widely in advising governments on the resource needs, expected impacts, and progression of HIV epidemics are compared in Table 2, including the AIDS Epidemic Model (AEM), the Estimation and Projections Package (EPP), the Goals Model, and the Modes of Transmission (MOT) model. Of these, only AEM and Goals are full process models that track epidemics over time, relate behavioral parameters to coverage, prevalence and incidence, and produce long-term forecasts. While AEM and Optima share a similar fundamental structure, AEM supports only a limited number of population groups; for example, transgender populations are important contributors to several national HIV epidemics (e.g., Indonesia), but there has historically been no scope for including this population group in AEM. AEM currently only has a basic costing component, which does not allow complete cost-effectiveness analyses nor resource estimation or allocation studies to be conducted. AEM has been used in a number of epidemic settings, particularly in Asia [13,25].

Table 2: Comparison of HIV epidemic model characteristics.

Model	Approach	Populations	Purpose	Inputs	Outputs
EPP	Fits four parameters to a simple model; written in Java	MSM, PWID, FSW, male SW, CSW, and low-risk (separated into urban and rural)	Estimate and project adult HIV prevalence and incidence	Size of subpopulations; HIV prevalence among subpopulations; treatment data	Current number of HIV infections; HIV infection trends (5-year projections)
AEM	Semi-empirical process model; written in Java	PWID, direct FSW, indirect FSW, MSW, CSW, and MSM	Provide a policy and planning tool for Asian countries	Size of subpopulations; HIV and STI prevalence; risk behavior data; average duration in each population	Trends of HIV infections; impacts on AIDS cases, ART needs, deaths, etc.

					(long-term projections)
MOT	Risk equations; written in Excel	PWID, FSW, MSM, and low-risk (separated into males and females)	Calculate expected number of infections over coming year	HIV prevalence; number of individuals with particular exposure; rates of exposure	Incidence (HIV acquisition) per risk group
Goals / Spectrum	Compartmental rate-based model; written in Visual Basic	MSM and high, medium, and low-risk groups	Estimate costs and impact of different interventions	Sexual behavior by risk group; demographic data; base year human capacity	Costs; HIV prevalence and incidence (5- year projections)
Optima	Compartmental rate-based model; versions available for MATLAB and Python	Flexible; unlimited but usually around 8-20 groups, including key affected and general populations and different age groups	Analyze and project HIV epidemics; determine optimal resource allocations	Size of population groups; HIV and STI prevalence; risk behavior data (e.g. condom use); biological constants (e.g. background death rates)	HIV prevalence and incidence trends; healthcare costs; deaths; optimal resource allocations

Abbreviations: EPP, Estimation & Projections Package; AEM, AIDS Epidemic Model; MOT, Modes of Transmission Spreadsheet; MSM, men who have sex with men; PWID, people who inject drugs; SW, sex workers; FSW, female sex workers; CSW, clients of sex workers; STI, sexually transmitted infections; ART, antiretroviral therapy.

EPP and Goals are both components of the Spectrum software package [11]. In contrast to AEM and Optima, EPP does not model the actual epidemic dynamics; instead, it is a phenomenological model that calibrates to past epidemic trends and projects them into the future. It has been used extensively to predict trends in African epidemics, and is used for HIV estimates in many countries around the world, including by UNAIDS [2,11,14,15].

The Goals model is closer to Optima in terms of its structure, data requirements, and purpose: both models require detailed demographic, epidemiological, behavioral, and clinical data, and both have been used as part of HIV strategic planning. The Goals model has been used by countries and UNAIDS in resource needs estimates and epidemic projections. Like Optima, the key inputs of the Goals model are used to determine the force-of-infection for each population. In addition, in the Goals model, both the population groups (including people who inject drugs, MSM, and low-, medium-, and high-risk individuals) and the types of HIV programs that can be modeled (including condom promotion programs, workplace programs, male circumcision, ART, HIV testing and counseling, and PMTCT) are fixed. In contrast, Optima allows HIV programs to be defined to target specific subpopulations within specific geographic areas, with different service delivery models, and with user-specified efficacies. Another difference between the Goals model and Optima is that the Goals model does not include undiagnosed people living with HIV, so HIV testing is instead assigned an HIV prevention benefit (i.e., HIV testing directly reduces HIV incidence due to behavioral changes by those who either test positive or negative).

MOT contains a detailed process model of HIV acquisition, requiring similar data inputs as are used in Optima to calculate the force-of-infection [16]. However, unlike Optima, AEM, and Goals, MOT's sole output is a prediction of the acquisition of new infections in each population group in the current year – which is one of the standard outputs produced by AEM, the Goals model, and Optima. Recently, concerns have been raised regarding the validity of the epidemic drivers identified by MOT [26].

Thus, while each of these previously used packages has a potentially useful role in informing policy-makers of certain aspects of an HIV epidemic, none of them allow a unified approach across (a) different types of epidemics (e.g., concentrated vs. generalized), (b) different types of

analysis (e.g., cost-effectiveness of past investments and projections of future epidemic trends at different funding levels and for different HIV program impact objectives), (c) different population groups (e.g., people who inject drugs, transgender individuals, migrant workers, and persons of different age groups), and (d) different HIV program impact targets at different funding levels.

Optima analyses translate inherent uncertainty in data and assumptions into uncertainty in outputs. The analyses also allow users to incorporate real-world constraints associated with all programs in the optimization analyses (e.g., no one who starts ART is to stop ART; programs cannot be immediately defunded but may only have reduced funding up to a certain percentage each year to enable a realistic transition). Graphical, tabular, and other numeric output provide stakeholders with clear qualitative and quantitative conclusions to assist in HIV policy and programming decision-making.

The main limitation of Optima is its requirement for data. Optima's strength is that it can utilize a far larger quantity of data than most other HIV models; however, this advantage wanes in the absence of good data availability. In general, the model used for a given analysis should be commensurate with the amount of data available: if ample data are available, a complex model such as Optima is likely to provide informative results; conversely, if data are strictly limited, then a simple model with few parameters (such as EPP) may give more robust and meaningful results on epidemic trajectories. However, the flexible nature of Optima means that by reducing the number of population groups and the number of modes of transmission, its data requirements can be reduced to be in line with those of other simpler models. Similarly, the large number of parameters supported by Optima risks over-parameterization and over-fitting. To circumvent this, most parameters in the model are set to values determined by the best available data; only a relatively small number of parameters (relating to the initial HIV prevalence, force-of-infection, and testing and treatment rates) are varied during calibration.

A second limitation of Optima is its reliance on assumed cost-coverage and coverage-outcome curves to determine the optimal allocations. Actual data from the field, along with consensus-deriving dialogue with country experts, can be used to inform the parameters of these curves and their uncertainty. Typically, the data available to constrain these may be limited, and thus assumptions must be made to fill these data gaps, particularly regarding expected behaviors for the extremes of zero and saturation funding. Since the optimal allocation is determined by the slopes of the cost-coverage-outcome relationships, the results are especially sensitive to uncertainty in these curves.

Third, since Optima is a population- and rate-based model, it relies on average quantities rather than full distributions. Thus, while it is possible to introduce heterogeneity in populations (e.g., population groups can be further subdivided into lower- and higher-risk subgroups), continuous distributions of risk behaviors are not modeled. In addition, rate-based models have inherent limitations in modeling certain kinds of processes. For example, while a death rate may be chosen that results in the correct mean life expectancy, a rate-based process will result in an exponential distribution of life expectancies, which is only a rough approximation of the true distribution, which is more Gaussian.

To allow policymakers to access Optima’s functionality without requiring detailed knowledge of computer programming or mathematical modeling, we have developed a Python-based version of the software, along with a user-friendly Javascript-based graphical interface (available via www.optimamodel.com). This version uses Amazon Web Services’ Elastic Compute Cloud [27,28] to implement parallelization, allowing the computationally intensive tasks of calibration, uncertainty analysis, and optimization to be performed with minimal delay to the user. Users are able to select and modify population groups and HIV programs to suit their needs, as well as define objectives and constraints relevant to their particular setting.

Optima has already been used in numerous countries to inform the development of HIV investment cases and concept notes (as are now required to apply for Global Fund grants), contribute to the development of National Strategic Plans and Operational Plans to help allocate domestic and international funds, and be used as an advocacy tool to demonstrate the benefits of past HIV investments and justify future resource allocations to HIV. In conclusion, we hope that Optima proves to be a useful tool to help policy-makers understand their countries’ HIV epidemics and to allocate resources efficiently for maximal impact.

Acknowledgements

This study and the Optima model was funded by the World Bank Group and the Australian National Health and Medical Research Council. The Kirby Institute at UNSW Australia is funded by the Australian Government, Department of Health and Ageing. The findings, interpretations, and conclusions expressed in this work are those of the authors and do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent.

References

1. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, et al. (2013) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* 380: 2197–2223.
2. Brown T, Grassly N, Garnett G, Stanecki K (2006) Improving projections at the country level: the UNAIDS Estimation and Projection Package 2005. *Sexually Transmitted Infections* 82: iii34–iii40.
3. Liu L, Johnson HL, Cousens S, Perin J, Scott S, et al. (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *The Lancet* 379: 2151–2161.
4. Vassall A, Remme M, Watts C, Hallett T, Siapka M, et al. (2013) Financing Essential HIV Services: A New Economic Agenda. *PLoS Medicine* 10: e1001567.
5. UNAIDS (2014) The Gap Report.
6. Beeharry G, Schwab N, Akhavan D, Hernandez R, Paredes C (2002) Optimizing the allocation of resources among HIV prevention interventions in Honduras: World Bank, Washington, DC.
7. Craig AP, Thein H-H, Zhang L, Gray RT, Henderson K, et al. (2014) Spending of HIV resources in Asia and Eastern Europe: systematic review reveals the need to shift funding allocations towards priority populations. *Journal of the International AIDS Society* 17: 18822.
8. Anderson S-J, Cherutich P, Kilonzo N, Cremin I, Fecht D, et al. (2014) Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: a modelling study. *The Lancet* 384: 249–256.

9. Eaton JW, Menzies NA, Stover J, Cambiano V, Chindelevitch L, et al. (2014) Health benefits, costs, and cost-effectiveness of earlier eligibility for adult antiretroviral therapy and expanded treatment coverage: a combined analysis of 12 mathematical models. *The Lancet Global Health* 2: e23-e34.
10. Naning H, Kerr C, Kamarulzaman A, Osornoprasop S, Dahlui M, et al. (2014) Return on Investment and Cost-Effectiveness of Harm Reduction Program in Malaysia.
11. Stover J, Brown T, Marston M (2012) Updates to the Spectrum/Estimation and Projection Package (EPP) model to estimate HIV trends for adults and children. *Sexually Transmitted Infections* 88: i11-i16.
12. Stover J, Johnson P, Zaba B, Zwahlen M, Dabis F, et al. (2008) The Spectrum projection package: improvements in estimating mortality, ART needs, PMTCT impact and uncertainty bounds. *Sexually Transmitted Infections* 84: i24-i30.
13. Brown T, Peerapatanapokin W (2004) The Asian Epidemic Model: a process model for exploring HIV policy and programme alternatives in Asia. *Sexually Transmitted Infections* 80: i19-i24.
14. Brown T, Bao L, Raftery AE, Salomon JA, Baggaley RF, et al. (2010) Modelling HIV epidemics in the antiretroviral era: the UNAIDS Estimation and Projection Package 2009. *Sexually Transmitted Infections* (Supplement).
15. Ghys PD, Brown T, Grassly N, Garnett G, Stanecki K, et al. (2004) The UNAIDS Estimation and Projection Package: a software package to estimate and project national HIV epidemics. *Sexually Transmitted Infections* 80: i5-i9.
16. Case KK, Ghys PD, Gouws E, Eaton JW, Borquez A, et al. (2012) Understanding the modes of transmission model of new HIV infection and its use in prevention planning. *Bulletin of the World Health Organization* 90: 831-838A.
17. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, et al. (2014) Estimating per-act HIV transmission risk: a systematic review. *AIDS* 28: 1509-1519.
18. Fox J, White PJ, Weber J, Garnett GP, Ward H, et al. (2011) Quantifying sexual exposure to HIV within an HIV-serodiscordant relationship: development of an algorithm. *AIDS* 25: 1065-1082.
19. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, et al. (2011) Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine* 365: 493-505.
20. Bonner K, Mezochow A, Roberts T, Ford N, Cohn J (2013) Viral load monitoring as a tool to reinforce adherence: a systematic review. *Journal of Acquired Immune Deficiency Syndromes* 64: 74-78.
21. Kerr C, Smolinski T, Dura-Bernal S, Wilson D (under review) Optimization by Bayesian adaptive locally linear stochastic descent. *Nature Scientific Reports*.
22. Gilks WR (2005) Markov Chain Monte Carlo: Wiley Online Library.
23. Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57: 97-109.
24. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21: 1087-1092.
25. Gouws E, White P, Stover J, Brown T (2006) Short term estimates of adult HIV incidence by mode of transmission: Kenya and Thailand as examples. *Sexually Transmitted Infections* 82: iii51-iii55.
26. Mishra S, Pickles M, Blanchard JF, Moses S, Shubber Z, et al. (2014) Validation of the Modes of Transmission model as a tool to prioritize HIV prevention targets: a comparative modelling analysis. *PLoS ONE* 9: e101690.
27. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, et al. (2010) A view of cloud computing. *Communications of the ACM* 53: 50-58.
28. Jackson KR, Ramakrishnan L, Muriki K, Canon S, Cholia S, et al. (2010) Performance analysis of high performance computing applications on the Amazon Web Services cloud. *IEEE Second International Conference on Cloud Computing Technology and Science*: 159-168.