Mathematical Biosciences and Engineering

MBE, 18(6): 7666–7684.
DOI: 10.3934/mbe.2021380
Received: 12 July 2021
Accepted: 25 August 2021
Published: 06 September 2021

http://www.aimspress.com/journal/MBE

Research article

A reinforcement learning model to inform optimal decision paths for HIV elimination

Seyedeh N. Khatami and Chaitra Gopalappa*

Mechanical and Industrial Engineering Department, University of Massachusetts Amherst, Amherst, MA 01003, USA

* **Correspondence:** Email: chaitrag@umass.edu; Tel: +1-4135452306.

Appendix

1. Depicting MDP as a decision tree

The sequential decision-making problem for HIV epidemic control and elimination can be visualized as a decision-tree. Figure A1 presents the decision tree for a simple 2-state system with n action choices, with the epidemic state in 2015 being the start node, actions being the decision nodes, epidemic states at 5-year intervals being the chance nodes, the possible epidemic states in 2070 being the end nodes, and a policy being a path in a decision-tree. The problem is to identify, from among all possible paths of a decision-tree, the one with the maximum total reward. The decision tree starts in the initial state S_0 and corresponds to the epidemic state in the year 2015 (t_0). Each square is a decision node at which there are multiple actions (stems branching out from a node) to choose from ($[a_1, a_2, \dots a_n]$ in Figure A1). An action taken in any year t_i transitions the epidemic to a different state in year t_{i+5}, with some uncertainty (oval) in the actual state it transitions to, denoted as probabilities $p_a(j,k)$ for transitioning to state S_k from state S_j when action a is taken. Thus, the decision tree branches out from the year 2015 to 2070, with decision-making occurring at every 5-year interval. The "total reward" is the output at the end of each branch in the year 2070. However, the problem cannot be solved using decision trees because it is impractical to evaluate all possible paths exhaustively. In this simple 2-state n-actions example, that would be $2n$ paths if evaluating up to time t_1, $(2n)^{2n}$ paths if evaluating up to time t_2, thus growing exponentially over the time-periods evaluated.

The HIV model introduced in this paper is significantly more extensive and has 16,501-states 36-actions evaluated for 11 time-periods. Hence, we formulated the problem as a Markov decision process (MDP), solved using optimization modeling that efficiently uses mathematical concepts to search through the solution space. MDP can also be solved by dynamic programming or reformulating as a linear programming optimization model; however, these methods also require the knowledge of the transition probability matrices (P_a) for each action a, square matrices with dimensions equal to the number of states ($p_a(j,k)$ are elements of P_a), which is infeasible for HIV. Therefore, to overcome these challenges, we use Q-learning with a stochastic dynamic simulation model to directly simulate the state transitions (replacing P_a) and estimate rewards.

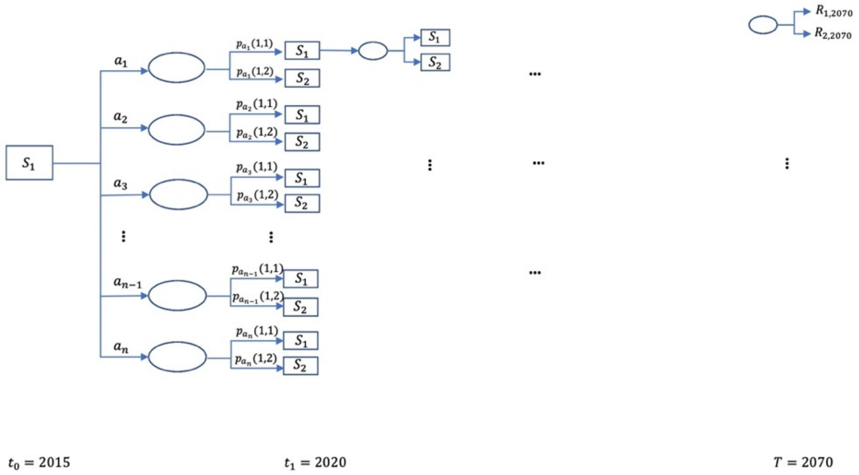


Figure A1. Decision tree representation of the Markov decision process problem (MDP); the objective of MDP is to find the path with an optimal trade-off in costs and benefits.

2. Estimating diagnostic and retention-in-care rates for a given action and simulating the action in PATH 2.0 model

As discussed in the main manuscript, we formulated an action as a combination of the percentage decrement in proportion unaware and the percentage increment in proportion on ART; each varying by transmission risk-group ($a_{unaware,i}, a_{ART,i}, \forall i \in \{HETs, MSM\}$). Corresponding to every action, we estimated the diagnostic rate (δ_t) and retention-in-care rate ($1-\rho_t$), and used the rates for simulating the action in the Progression and Transmission of HIV (PATH 2.0) model , including estimating the number of persons successfully intervened (newly diagnosed and retained-in-care) and the corresponding intervention costs. The estimation method includes a combination of expressing the system as a differential equations model and utilizing the PATH 2.0 model. In this section, we present the method of estimating δ_t and $1-\rho_t$ using $a_{unaware,i}$ and $a_{ART,i}$, the corresponding number of persons successfully intervened (newly diagnosed and retained-in-care), and simulating the action in PATH 2.0. In addition, we present the estimation of the intervention costs of that action in Appendix Section 4.

We can express the disease incidence and transition along the stages of care-continuum of HIV-infected persons as a set of differential equations or a compartmental model, with each disease and care continuum stage representing a compartment (Figure A2). We have four compartments, susceptible (S), infected and unaware (U), infected and aware but not in care (NC), and infected in care and on ART (ART).

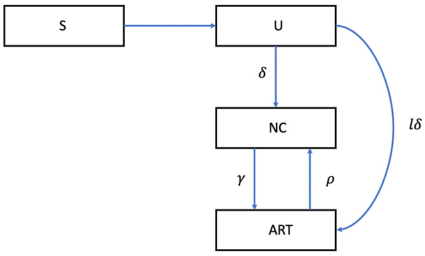


Figure A2. Flow diagram for disease incidence and transitioning along the stages of care-continuum HIV infected persons. S: susceptible, U: infected and unaware, NC: infected and aware but not in care, ART: infected, in care and on ART, δ: diagnostic rate, γ: rate of entering care and treatment among those not in care, and ρ: rate of dropping-out of care, and lδ: rate of diagnosis and linked to care.

Let,

I_t be the number of infected persons at time t,

i_t be the number of new infections at time t,

$p_{t,s}$ be the proportion of infected persons in stage s $\in \{\mathrm{U}, \mathrm{NC}, \mathrm{ART}\}$, such that $p_{t,s}I_t$, would be the number of infected people in stage s at time t,

δ_t be the diagnostic rate at time t,

l_t be the proportion of persons linking to care and initiating treatment upon diagnosis, among those diagnosed at time t ,

γ_t be the rate of entering care and treatment among those not in care at time t,

ρ_t be the rate of dropping-out of care and treatment at time t, and

m_t be the number of new deaths at time t.

Then, at a sufficiently small incremental time-step $\mathrm{t}+1$ (we use monthly increments), we can write the equations for the number of people in each stage by formulating as a system of differential equations,

$$p_{t+1,s}I_{t+1} = p_{t,s}I_t + \frac{dp_{t,s}I_t}{dt} \tag{1}$$

where, $\frac{dp_{t,s}I_t}{dt}$ is the rate of the change in $p_{t,s}I_t$, i.e., the change in the number of infected persons in stage s at t.

Specifically, for each stage s $\in \{\mathrm{U}, \mathrm{NC}, \mathrm{ART}\}$ we can apply Eq (1) and write,

$$I_t p_{t,U} = I_{t-1}p_{t-1,U} + i_t - \delta_t I_{t-1} p_{t-1,U} \tag{2}$$

$$I_t p_{t,NC} = I_{t-1}p_{t-1,NC} + \delta_t(1-l_t)I_{t-1}p_{t-1,U} - \gamma_t I_{t-1}p_{t-1,NC} + \rho_t I_{t-1}p_{t-1,ART} \tag{3}$$

$$I_t p_{t,ART} = I_{t-1} p_{t-1,ART} + \delta_t l_t I_{t-1} p_{t-1,U} + \gamma_t I_{t-1} p_{t-1,NC} - \rho_t I_{t-1} p_{t-1,ART} \tag{4}$$

Additionally, as $p_{t,s}$ is the 'proportion' of infected people in stage s, summing over all disease stages should add to 1, or

$$p_{t,NC} = 1 - \left(p_{t,U} + p_{t,ART}\right) \tag{5}$$

Further, the number of infected people at t (I_t) would be the number infected at $\mathrm{t} - 1$ plus new infections (i_t) minus deaths (m_t), i.e.,

$$I_t = I_{t-1} + i_t - m_t \tag{6}$$

Note, these equations are applied for each risk group (heterosexuals, and MSM) separately, but we do not indicate the risk group in subscripts for clarity of notations.

Estimate diagnostic rate ($\boldsymbol{\delta_t}$) and the corresponding number of persons successfully intervened (newly diagnosed)

We can write the monthly change in proportion unaware as $-\left(p_{t-1,U} - p_{t,U}\right) = \hat{a}_{unaware,t}$, which is the selected proxy action choice for the change in proportion unaware at decision epoch T divided by 60 months ($\text{i.e.}, \hat{a}_{unaware,t} \in \frac{a_{unaware,i}}{60}$), as our decision-making interval is every five years. Note that $\hat{a}_{unaware,t}$ is specific to risk-group i but we drop subscript i for clarity of notation. We calculate monthly estimates as the time-step in the PATH 2.0 simulation model is monthly. We assume the changes in proportion unaware are achieved uniformly over the 60-month period, and thus, the estimated diagnostic rate would be representative of linearly scaling up HIV-testing interventions over this interval. Then, using Eq (6), we can express diagnostic rate by rewriting Eq (2) as

$$\delta_t = \frac{-(I_{t-1} + i_t - m_t)p_{t,U} + I_{t-1}p_{t-1,U} + i_t}{I_{t-1}p_{t-1,U}}. \tag{7}$$

By substituting $p_{t,U} = p_{t-1,U} + \hat{a}_{unaware,t}$ in Eq (7) we can write

$$\delta_t = \frac{-(I_{t-1} + i_t - m_t)\left(p_{t-1,U} + \hat{a}_{t,unaware}\right) + I_{t-1}p_{t-1,U} + i_t}{I_{t-1}p_{t-1,U}} \tag{8}$$

and the corresponding number of persons to diagnose as

$$\delta_t I_{t-1} p_{t-1,U} = -(I_{t-1} + i_t - m_t)\left(p_{t-1,U} + \hat{a}_{t,unaware}\right) + I_{t-1}p_{t-1,U} + i_t. \tag{9}$$

Estimate retention-in-care rate and the number of persons successfully intervened (retained-in-care)

We can write the monthly change in proportion on ART as $\left(p_{t,ART} - p_{t-1,ART}\right) = \hat{a}_{ART,t}$, which is the selected proxy action choice for the change in proportion on ART at decision epoch T divided by 60 months ($\text{i.e.}, \hat{a}_{ART,t} \in \frac{a_{ART,i}}{60}$), as our decision-making interval is every five years. Note that, $\hat{a}_{ART,t}$ is specific to risk-group i, but we drop subscript i for clarity of notation. We assume the changes in proportion on ART are achieved uniformly over the 60 month period, and thus, the estimated retention-in-care rates would be representative of linearly scaling up retention-in-care interventions over this interval. We set $p_{t,ART} = p_{t-1,ART} + \hat{a}_{ART,t}$ in Eq (4) and rewrite it to express the rate of dropping-out as

$$I_t p_{t,ART} = I_{t-1}p_{t-1,ART} + \delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1}p_{t-1,NC} - \rho_t I_{t-1}p_{t-1,ART},$$

$$\rho_t = \frac{-(I_{t-1} + i_t - m_t)p_{t,ART} + I_{t-1}p_{t-1,ART} + \delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1}p_{t-1,NC}}{I_{t-1}p_{t-1,C}},$$

$$\rho_t = \frac{-(I_{t-1} + i_t - m_t)\left(p_{t-1,ART} + \hat{a}_{ART,t}\right) + I_{t-1}p_{t-1,ART} + \delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1}p_{t-1,NC}}{I_{t-1}p_{t-1,ART}}, \tag{10}$$

and estimate the corresponding number of persons dropping-out of care as

$$\rho_t I_{t-1}p_{t-1,ART} = -(I_{t-1} + i_t - m_t)\left(p_{t-1,ART} + \hat{a}_{ART,t}\right) + I_{t-1}p_{t-1,ART} + \delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1}p_{t-1,NC}. \tag{11}$$

Following from above, we estimate the number of persons retained in care as the number of persons not dropping-out of care, i.e., $(1 - \rho_t)I_{t-1}p_{t-1,ART}$.

Simulating the action in PATH 2.0

Simulating any given action $a_{unaware,i}, a_{ART,i}, \forall i \in \{HETs, MSM\}$ over the 5-year interval (in monthly time-steps) involves iteratively simulating PATH every month for estimating the terms on the right-hand side of Eqs (8) and (10), solving for diagnostic rate (δ_t) and retention-in-care rate $(1 - \rho_t)$ for that month using (8) and (10), estimating the numbers to diagnose and retain-in-care using Eqs (9) and (11), and using that in PATH to simulate diagnosis and care events for that month. Details of the PATH 2.0 model along with the estimation of the parameters in Eqs (8) and (10) and simulating care and diagnosis events are discussed elsewhere , here we only give a brief description related to the estimation of parameters in Eqs (8) and (10). PATH 2.0 model was initialized to be representative of the HIV infected population in the US in 2006, using data from the US National HIV Surveillance System (NHSS), and simulated to 2015. It keeps track of I_t and $p_{t,s}$ over time. It estimates new infections (i_t) by modeling transmissions at the individual-level for every susceptible-infected partnership, modeling it as a function of stage $\mathrm{s} \in \{\mathrm{U, NC, ART}\}$ of the infected person at time $t - 1$. Thus, the number of new infections is a function of $p_{t-1,U}$ and $p_{t-1,NC}$. It estimates the number of deaths ($\mathrm{m_t}$) by simulating mortality at the individual-level for each person using stage- and age-specific mortality rates. It uses annual data from NHSS for the proportion linking to care (l_t) upon diagnosis and assume it will be maintained at the 2015 level for future years . It uses re-entry rates (γ_t) from studies in the literature .

3. Mathematical advantage of formulating action space as changes in proportions unaware and on ART instead of diagnostic and retention-in-care rates

As discussed in section 2, we formulate the action space (A) as a combination of changes in proportions unaware and on ART, i.e., $\mathrm{A} = \left\{\left[a_{unaware,i}, a_{ART,i}, \forall i \in \{HETs, MSM\}\right]\right\}$, as proxy for diagnostic and retention-in-care rates $\{[\delta_i, 1-\rho_i, \forall i \in \{HETs, MSM\}]\}$. We prove here that the proxy metrics efficiently constrain the size of the action space, which would increase the chance of convergence of the Q-learning algorithm. It is also computationally efficient as it requires fewer evaluations of the simulation model. We also prove that formulating an action as $[\Delta\delta_i, 1-\Delta\rho_i, \forall i \in \{HETs, MSM\}]$ also leads to a large action space, where $\Delta\delta_i$ and $1-\Delta\rho_i$ are changes in δ_i and ρ_i, respectively, over two consecutive decision epochs, as the proxies $a_{unaware,i}$ and $a_{ART,i}$ are also decrements or increments (of $\mu_{u,i}$ or $\mu_{v,i}$, respectively) over two consecutive decision epochs. We also prove that, corresponding to every combination of $\left[\hat{a}_{unaware,t}, \hat{a}_{ART,t}\right]$ there is a unique combination of $[\delta_t, 1-\rho_t]$, and thus solving for the optimal combination of the proxy metrics is equivalent to solving for the optimal diagnostic and retention-in-care rates. We prove these through Remarks 1 to 4.

Note that, $\left[\hat{a}_{unaware,t}, \hat{a}_{ART,t}\right]$ and $[\delta_t, 1-\rho_t]$ have t (time) subscripts as they are the monthly values defined in Appendix Section 2 corresponding to a proxy action $\left[a_{unaware,i}, a_{ART,i}, \forall i\right]$ and original action $\{[\delta_i, 1-\rho_i, \forall i]\}$, respectively. They do vary by risk-group but we drop the subscripts i for clarity of notations. Without loss of generality, we use $\left[\hat{a}_{unaware,t}, \hat{a}_{ART,t}\right]$ and $[\delta_t, 1-\rho_t]$ to prove our Remarks about $\left[a_{unaware,i}, a_{ART,i}, \forall i \in \{HETs, MSM\}\right]$ and $[\delta_i, 1-\rho_i, \forall i \in \{HETs, MSM\}]$, respectively.

Remark 1: Given the system state x at time $\mathrm{t}-1$, $(X_{t-1} = x)$, corresponding to every action, $a_{unaware,i}$, there is a unique diagnostic rate, δ_t, i.e., $f\colon \hat{a}_{unaware,t} \rightarrow \delta_t$ is a bijection function and corresponding to every action, $a_{ART,i}$, there is a unique retention-in-care rate, $1-\rho_t$, i.e., $g\colon \hat{a}_{ART,t} \rightarrow 1-\rho_t$ is a bijection function.

Proof: From Eq (8) we have diagnostic rate as:

$$\delta_t = \frac{-(I_{t-1} + i_t - m_t)\left(p_{t-1,U} + \hat{a}_{unaware,t}\right) + I_{t-1}p_{t-1,U} + i_t}{I_{t-1}p_{t-1,U}},$$

the only controllable unknown is $\hat{a}_{unaware,t}$ as all other parameters correspond to or are calculated using system state at time $t-1$, as discussed in Section 2. Therefore, δ_t is a linear function of $\hat{a}_{unaware,t}$, i.e., $f\colon \hat{a}_{unaware,t} \rightarrow \delta_t$ is a bijection function. This implies that, at any given system state at time $t-1$ for every action $\hat{a}_{unaware,t}$, we can calculate a unique value for the diagnostic rate.

Similarly, from Eq (10) we have drop-out rate as:

$$\rho_t = \frac{-(I_{t-1} + i_t - m_t)\left(p_{t-1,ART} + \hat{a}_{ART,t}\right) + I_{t-1}p_{t-1,ART} + \delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1}p_{t-1,NC}}{I_{t-1}p_{t-1,ART}}$$

the only controllable unknown is $\hat{a}_{ART,t}$ as all other parameters correspond to or are calculated using system state at time $t-1$, as discussed earlier in Section 2. Therefore, ρ_t is a linear function of $\hat{a}_{ART,t}$, i.e., $g\colon \hat{a}_{ART,t} \rightarrow \rho_t$ is a bijection function. This implies that, at any given system state at time $t-1$, for every action a_{ART}, we can calculate a unique value for the retention-in-care rate $1-\rho_t$.

Remark 2: From a public health perspective, all actions that result in $\left(p_{t-1,U} - p_{t,U}\right) < 0$ or $\left(p_{t,ART} - p_{t-1,ART}\right) < 0$ are undesirable and should not be selected.

Proof: If the conditions are true, it would imply that a larger proportion of people with HIV are unaware of their infection and/or are not on treatment at time t compared to $\mathrm{t} - 1$. Being unaware and not on treatment are associated with an increase in transmissions and mortalities. Thus, all actions that result in $\left(p_{t-1,U} - p_{t,U}\right) < 0$ or $\left(p_{t,ART} - p_{t-1,ART}\right) < 0$ are undesirable as they worsen the epidemic and should not be selected.

Remark 3: Setting action space as $\mathrm{A} = \{[a_{unaware,i}, a_{ART,i}, \forall i]\}$ instead of $\mathrm{A} = \{[\delta_i, 1 - \rho_i, \forall i]\}$ efficiently controls the number of possible interventions and thus is more computationally efficient.

Proof: If action space $\mathrm{A} = \{[a_{\text{unaware}}, a_{ART}]\}$:

As $\hat{a}_{unaware,t} = -\left(p_{t-1,U} - p_{t,U}\right)$, we can directly select actions such that $-\left(p_{t-1,U} - p_{t,U}\right) \geq 0$ (see Remark 2). As we use a_{unaware} to estimate diagnostic rate, it constrains diagnostic rates to only those that correspond to desirable outcomes. Thus, all testing intervention programs that are below the minimum diagnostic rates can be excluded. The selection of action is not dependent on any parameters of the system, except for the feasibility constraint $\mathrm{p_{t,U}} > 10\%$, as discussed in Section 2.1 of the manuscript.

Similarly, as $a_{ART} = -\left(p_{t-1,ART} - p_{t,ART}\right)$, we can directly select actions such that $\left(p_{t,ART} - p_{t-1,ART}\right) \geq 0$ (see Remark 2), naturally constraining the selection of retention-in-care rates to those that would result in desirable outcomes. Thus, all intervention programs whose efficacy is below the minimum retention-in-care rates estimated here can be excluded. The selection of action is not dependent on any parameters of the system, except for the feasibility constraint $p_{t,ART} < 90\%$, as discussed in the Section 2.1.

Rearranging Eq (7) for diagnostic rate from above, we can write

$$\delta_t = \frac{-(i_t - m_t)p_{t,U} + I_{t-1}\left(p_{t-1,U} - p_{t,U}\right) + i_t}{I_{t-1}p_{t-1,U}} \tag{12}$$

And further rearranging to write

$$\left(p_{t-1,U} - p_{t,U}\right) = \frac{\delta_t I_{t-1} p_{t-1,U} - \left[i_t - (i_t - m_t)p_{t,U}\right]}{I_{t-1}} \tag{13}$$

As $p_{t,U} \leq 1$, if $i_t > m_t$ then $-(i_t - m_t)p_{t,U} < i_t$; and if $i_t < m_t$ then $-(i_t - m_t)p_{t,U} > 0$. Therefore, the following condition is always true $[i_t - (i_t - m_t)p_{t,U}] > 0$.

The above implies that, there are certain values of δ_t such that $\delta_t I_{t-1} p_{t-1,U} \geq \left[i_t - (i_t - m_t)p_{t,U}\right]$, which would yield $\left(p_{t-1,U} - p_{t,U}\right) \geq 0$, and certain other values of δ_t such that $\delta_t I_{t-1} p_{t-1,U} < \left[i_t - (i_t - m_t)p_{t,U}\right]$, which would yield $\left(p_{t-1,U} - p_{t,U}\right) < 0$, which is an undesirable outcome from a public health perspective. As the values of δ_t that generate $\delta_t I_{t-1} p_{t-1,U} \geq$

$[i_t - (i_t - m_t)p_{t,U}]$ is time-dependent on values of I_{t-1}, $p_{t-1,U}$, i_t, and m_t, a large set of values for δ_t should be evaluated as part of the action space. This is computationally expensive, and moreover inefficient, as many cases would result in an undesirable outcome.

Rearranging the equation for drop-out rate from Eq (10), we can write

$$\rho_t = \frac{-(I_{t-1} + i_t - m_t)p_{t,ART} + I_{t-1}p_{t-1,ART} + \delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1}p_{t-1,NC}}{I_{t-1}p_{t-1,ART}} \tag{14}$$

$$\rho_t = \frac{-(i_t - m_t)p_{t,ART} - I_{t-1}(p_{t,ART} - p_{t-1,ART}) + \delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1}p_{t-1,NC}}{I_{t-1}p_{t-1,ART}} \tag{15}$$

$$(p_{t,ART} - p_{t-1,ART}) = \frac{-I_{t-1}\rho_t p_{t-1,ART} - (i_t - m_t)p_{t,ART} + \delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1}}{I_{t-1}} \tag{16}$$

If $i_t > m_t$ then $(i_t - m_t)p_{t,ART} > 0$ and as $+\delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1} > 0$ and $-I_{t-1}\rho_t p_{t-1,ART} < 0$, there are certain combinations that can result $(p_{t,ART} - p_{t-1,ART})$ <0.

Similarly, if $i_t < m_t$ then $(i_t - m_t)p_{t,ART} < 0$ and as $+\delta_t l_t I_{t-1}p_{t-1,U} + \gamma_t I_{t-1} > 0$ and $-I_{t-1}\rho_t p_{t-1,ART} < 0$, there are certain combinations that can result in $(p_{t,ART} - p_{t-1,ART})$ <0.

As the values of ρ_t that result in $(p_{t,ART} - p_{t-1,ART})$ <0 are time-dependent on values of I_{t-1}, $p_{t-1,ART}$, i_t, and m_t a large set of values for ρ_t and thus $1 - \rho_t$ should be part of the action space, which is again computationally expensive and inefficient as many cases would result in undesirable outcomes.

Remark 4: Setting action space as $A = \{[a_{unaware,i}, a_{ART,i}, \forall i]\}$ instead of $A = \{[\Delta\delta_i, 1 - \Delta\rho_i, \forall i]\}$ efficiently controls the number of possible interventions and thus is more computationally efficient.

Proof: If action space $A = \{[a_{unaware}, a_{ART}]\}$: We discuss this case in Remark 3.

If action space $A = \{[\Delta\delta_i, 1 - \Delta\rho_i, \forall i]\}$ the corresponding rates to simulate at every time-step would be $\delta_t - \delta_{t+1}$ and $\rho_t - \rho_{t+1}$:

Without loss of generality, we prove this Remark by showing that to generate one scenario equivalent of $(p_{t-1,U} - p_{t,U}) = (p_{t,U} - p_{t+1,U}) = 0$, i.e., keeping proportion unaware constant over two consecutive time-steps, it would require evaluations of multiple combinations of δ_t and δ_{t+1} as the combination that generates $(p_{t-1,U} - p_{t,U}) = (p_{t,U} - p_{t+1,U}) = 0$ would be dependent on system parameters at that time.

Writing equations for δ_t and δ_{t+1}, setting $(p_{t-1,U} - p_{t,U}) = (p_{t,U} - p_{t+1,U}) = 0$, and subtracting one from the other, we get

$$\delta_t = \frac{-(i_t - m_t)p_{t,U} + i_t}{I_{t-1}p_{t-1,U}} \tag{17}$$

$$\delta_{t+1} = \frac{-(i_{t+1} - m_{t+1})p_{t+1,U} + i_{t+1}}{I_t p_{t,U}} \tag{18}$$

Subtracting Eqs (17) and (18) we get:

$$\delta_t - \delta_{t+1} = \frac{-(i_t - m_t)p_{t,U} + i_t}{I_{t-1}p_{t-1,U}} - \frac{-(i_{t+1} - m_{t+1})p_{t+1,U} + i_{t+1}}{I_t p_{t,U}} \tag{19}$$

As $\left(p_{t-1,U} - p_{t,U}\right) = 0$ and $\left(p_{t,U} - p_{t+1,U}\right) = 0$, we can set $p_{t-1,U} = p_{t,U} = p_{t+1,U}$:

$$\mathrm{\delta_t - \delta_{t+1} = \frac{-(i_t - m_t)p_{t,U} + i_t}{I_{t-1}p_{t,U}} - \frac{-(i_{t+1} - m_{t+1})p_{t,U} + i_{t+1}}{I_t p_{t,U}}} \tag{20}$$

$$\mathrm{\delta_t - \delta_{t+1} = \frac{I_t m_t p_{t,U} + I_t i_t\left(1 - p_{t,U}\right) - I_{t-1} m_{t+1} p_{t,U} - I_{t-1} i_{t+1}\left(1 - p_{t,U}\right)}{I_t I_{t-1} p_{t,U}}} \tag{21}$$

$$\mathrm{\delta_t - \delta_{t+1} = \frac{(I_t m_t - I_{t-1} m_{t+1})p_{t,U} + (I_t i_t - I_{t-1} i_{t+1})\left(1 - p_{t,U}\right)}{I_t I_{t-1} p_{t,U}}}. \tag{22}$$

Therefore, $\left(p_{t-1,U} - p_{t,U}\right) = \left(p_{t,U} - p_{t+1,U}\right) = 0$ could result from a range of diagnostic rate changes, $\delta_t - \delta_{t+1} > 0$, < 0, or $= 0$ depending on values of $\mathrm{I_t, m_t, I_{t-1}, m_{t+1}, i_t}$, and $\mathrm{i_{t+1}}$. This implies that, if formulating an action as $[\Delta\delta_i, 1 - \Delta\rho_i, \forall i]$, a large subset of values for $\delta_t - \delta_{t+1}$ should be evaluated as part of the action space. This is computationally expensive and, moreover, inefficient as many cases would result in an undesirable outcome.

Similarly, without loss of generality, to generate one scenario equivalent of $\left(p_{t-1,ART} - p_{t,ART}\right) = \left(p_{t,ART} - p_{t+1,ART}\right) = 0$, i.e., keeping proportion on ART constant over two consecutive time-steps, it would require evaluations of multiple combinations of ρ_t and ρ_{t+1} as the combination that generates $\left(p_{t-1,ART} - p_{t,ART}\right) = \left(p_{t,ART} - p_{t+1,ART}\right) = 0$ would be dependent on system parameters at that time.

Writing equations for ρ_t and ρ_{t+1}, setting $\left(p_{t-1,ART} - p_{t,ART}\right) = \left(p_{t,ART} - p_{t+1,ART}\right) = 0$, and subtracting one from the other, we get

$$\mathrm{\rho_t = \frac{-(i_t - m_t)p_{t,ART} + \delta_t l_t I_{t-1} p_{t-1,U} + \gamma_t I_{t-1} p_{t-1,NC}}{I_{t-1} p_{t-1,ART}}} \tag{23}$$

$$\mathrm{\rho_{t+1} = \frac{-(i_{t+1} - m_{t+1})p_{t+1,V} + \delta_{t+1} l_{t+1} I_t p_{t,U} + \gamma_{t+1} I_t p_{t,NC}}{I_t p_{t,ART}}} \tag{24}$$

$$\rho_t - \rho_{t+1} = \frac{-(i_t - m_t)p_{t,ART} + \delta_t l_t I_{t-1} p_{t-1,U} + \gamma_t I_{t-1} p_{t-1,NC}}{I_{t-1} p_{t-1,ART}} - \frac{-(i_{t+1} - m_{t+1})p_{t+1,ART} + \delta_{t+1} l_{t+1} I_t p_{t,U} + \gamma_{t+1} I_t p_{t,NC}}{I_t p_{t,ART}} \tag{25}$$

As $\left(p_{t-1,ART} - p_{t,ART}\right) = 0$ and $\left(p_{t,ART} - p_{t+1,ART}\right) = 0$, we can set

$$p_{t-1,ART} = p_{t,ART} = p_{t+1,ART}:$$

$$\rho_t - \rho_{t+1} = \frac{-(i_t - m_t)p_{t,ART} + \delta_t l_t I_{t-1} p_{t-1,U} + \gamma_t I_{t-1} p_{t-1,NC}}{I_{t-1} p_{t,ART}} - \frac{-(i_{t+1} - m_{t+1})p_{t,ART} + \delta_{t+1} l_{t+1} I_t p_{t,U} + \gamma_{t+1} I_t p_{t,NC}}{I_t p_{t,ART}} \tag{26}$$

Rewriting with a common denominator,

$$\rho_t - \rho_{t+1} = \frac{\begin{bmatrix} I_t m_t p_{t,ART} - I_t i_t p_{t,ART} + I_{t-1} i_{t+1} p_{t,ART} - I_{t-1} m_{t+1} p_{t,ART} \\ + I_t\left(\delta_t l_t I_{t-1} p_{t-1,U} + \gamma_t I_{t-1} p_{t-1,NC}\right) - I_{t-1}\left(\delta_{t+1} l_{t+1} I_t p_{t,U} + \gamma_{t+1} I_t p_{t,NC}\right) \end{bmatrix}}{I_t I_{t-1} p_{t,ART}} \tag{27}$$

Rearranging the numerator,

$$\rho_t - \rho_{t+1} = \frac{\begin{bmatrix} [I_t(m_t - i_t) + I_{t-1}(i_{t+1} - m_{t+1})]p_{t,ART} \\ + I_t\left(\delta_t l_t I_{t-1} p_{t-1,U} + \gamma_t I_{t-1} p_{t-1,NC}\right) - I_{t-1}(\delta_{t+1} l_{t+1} I_t p_{t,U} + \gamma_{t+1} I_t p_{t,NC}) \end{bmatrix}}{I_t I_{t-1} p_{t,ART}} \tag{28}$$

Therefore,$\left(p_{t-1,ART} - p_{t,ART}\right) = \left(p_{t,ART} - p_{t+1,ART}\right) = 0$ could result from a range of drop-out rate changes, $\rho_t - \rho_{t+1} > 0, < 0$, or $= 0$, depending on values of $I_t, m_t, I_{t-1}, m_{t+1}, i_t,$ and i_{t+1}. This implies, if formulation an action as $[\Delta\delta_i, 1 - \Delta\rho_i, \forall i]$, a large subset of values for $\rho_t - \rho_{t+1}$ should be evaluated as part of the action space. This is computationally expensive, and moreover inefficient, as most of the values would result in an undesirable outcome.

From Remarks 3 and 4, we can conclude that formulating the action space as $\mathrm{A} = \{[\delta_\mathrm{i}, 1 - \rho_\mathrm{i}, \forall \mathrm{i} \in \{\mathrm{HETs}, \mathrm{MSM}\}]\}$ would require evaluations of a large set of actions. For instance, if we use changes in testing and retention-in-care rates, we should consider both increase and decrease in these rates as both combinations can lead to higher proportion aware and proportion on ART. In this approach, choices would be combination of -25, 0, and 25% for testing rate and -20, -10, 0, 10, and 20% for retention-in-care (3 x 5 for HETs x 3 x 5 for MSM = 225). Or if we use changes in testing frequency and changes in retention-in-care rate as the action space, choices could be to test every 1, 2, 3, …, 10 years and retention-in-care rate can change by -20, -10, 0, 10, and 20% (10 x 5 for HETs x 10 x 5 for MSM = 2500). Problems with large action space generate issues of convergence, becoming infeasible to solve, and moreover, in this case, computationally inefficient, as a considerable portion of those values would result in $\left(p_{t-1,U} - p_{t,U}\right) < 0$, which is an undesirable outcome from a public health perspective (Remark 2). On the contrary, using action space $\mathrm{A} = \left\{\left[a_{\mathrm{unaware,i}}, a_{ART,i} \forall i\right]\right\}$ would naturally constrain the action space (to size 36 as discussed in paper section 2.1) by removing those actions that would result in undesirable outcomes, and is thus more efficient. Further, Remark 1 concludes that, for any given system state, solving for the optimal action $\left[a_{\mathrm{unaware,i}}, a_{ART,i} \forall i\right]$ is equivalent to solving for optimal $[\delta_\mathrm{i}, 1 - \rho_\mathrm{i}, \forall \mathrm{i} \in \{\mathrm{HETs}, \mathrm{MSM}\}]$.

4. Cost functions

We estimated the total cost of an action as the summation of the corresponding cost of the testing intervention program, retention-in-care intervention program, and treatment. The treatment costs are estimated in the PATH 2.0 simulation model by applying regimen-specific costs at the individual-level and are discussed elsewhere . In section 2, for every action, we estimated the number of persons successfully intervened (numbers newly diagnosed and retained-in-care). Corresponding to those numbers, we discuss the estimation of the corresponding costs of testing and retention-in-care intervention programs in this section.

a. Estimation of HIV testing costs

In the estimation of testing costs, we make the following assumptions based on currently available data on testing behavior and testing intervention programs [4,5,6]. HIV testing programs can be conducted in clinical or non-clinical settings, each having its own fixed and variable costs . Fixed cost includes the cost of clinics, other infrastructure, devices, equipment, staff, etc., while variable cost includes the cost per person tested. The marginal variable cost per additional person tested is a non-linear function of the proportion of the population tested and is influenced by the type of outreach program needed. Some people get tested voluntarily and incur only the cost of testing, while some get tested as a result of implementing an outreach intervention and thus incur additional costs of intervention . Outreach intervention can include providers reaching out to the client's community, delivering health information, reaching populations who have not been part of the traditional healthcare delivery system, HIV awareness campaigns, etc.. The outreach intervention program is not 100% effective, meaning that not all outreached persons would get tested for HIV. To achieve the required number of persons tested under any action, outreach programs maybe be necessary. Under any given system state for proportion unaware, the corresponding cost of outreach is a non-linear function of the number of people outreached, i.e., the marginal cost to achieve one additional HIV-positive test increases as the proportion unaware decreases indicating more efforts would be needed to reach a larger portion of the population . The cost per person for persons testing positive is different from the cost per person for persons testing negative, as persons testing positive also undergo follow-up confirmation tests and additional care services for linkage to care . In accordance with current CDC recommendations, we assumed only persons with high risk are recommended for regular testing and applied testing costs for only these populations. We assumed that 6% of heterosexual females, 10% of heterosexual males, and all MSM are high-risk .

We estimate the cost of testing corresponding to an action a as follows:

Let

$r_{t,a}$ be the number of persons testing positive at time step t under action a ,

$n_{t,a}$ be the number of persons testing negative at time step t under action a,

$x_{t,a}$ be the number of persons reached through an outreach testing intervention program at time step t under action a,

Y_v be the variable cost per person testing positive,

X_v be the variable cost per person testing negative,

O_v be the variable cost per person outreached through an outreach testing intervention program,

$X_{f-cl,a}$ be the total fixed cost of testing in a clinical setting under action a,

$X_{f-ncl,a}$ be the total fixed cost of testing in a non-clinical setting under action a,

$X_{f-o,a}$ be the total fixed cost of implementing an outreach intervention program under action a, and

C_{a_t} be the total cost of testing under action a.

Then, we can calculate the total cost of testing action as

$$C_{a_t} = r_{t,a}Y_v + x_{t,a}O_v + n_{t,a}X_v + X_{f-cl,a} + X_{f-ncl,a} + X_{f-o,a}. \tag{29}$$

Note that $r_{t,a}$ is the number of persons successfully intervened, i.e., the persons newly diagnosed as estimated in Appendix section 2. To achieve this, a larger number of persons must have tested negative ($n_{t,a}$), and further, a larger number of persons must have been reached through an intervention ($x_{t,a}$) as it is not necessary that all who are intervened would take a HIV test. Therefore, using our estimates of $r_{t,a}$ and based on other data from the literature on the effectiveness of interventions programs, we estimate $n_{t,a}$ and $x_{t,a}$. Furthermore, as the unit costs for persons in each category ($r_{t,a}$, $n_{t,a}$, or $x_{t,a}$) are also likely different, we estimate the unit costs for each category (Y_v, O_v, X_v, respectively) in addition to the fixed costs using data from the literature. Below, we discuss the estimation of each of these components on the right-hand side of the (29).

Estimation of $r_{t,a}$- Number of persons testing positive at time step t under action a:
Note that this is the number of persons successfully intervened, i.e., the persons newly diagnosed as estimated in Appendix section 2, i.e.,

$$r_{t,a} = \delta_t I_{t-1} p_{t-1,U} \tag{30}$$

where

δ_t is the rate of HIV diagnosis under action a,
I_t is the number of infected persons at time t, and
$p_{t,U}$ is the proportion of infected persons in stage unaware at time *t*.
Estimation of δ_t was discussed earlier, and I_{t-1} and $p_{t-1,U}$ are simulated in the PATH 2.0 model.

Estimation of $x_{t,a}$ - Number of people outreached at time step t under action a:

Let
μ be the rate of diagnosis through voluntary testing,
θ be the effectiveness of an outreach intervention program,
φ be the proportion of people testing positive for HIV among those tested through an outreach intervention program, and
N_t be the total population at time *t*.

Then we can write,
$\mu I_{t-1} p_{t-1,U}$ as the number of persons diagnosed through voluntary testing, and
$x_{t,a}\theta\varphi$ as the number of persons diagnosed through an outreach testing intervention program, i.e.,

$$x_{t,a}\theta\varphi = (\delta_t I_{t-1} p_{t-1,U}) - \left(\mu I_{t-1} p_{t-1,U}\right) \tag{31}$$

And thus, estimate the total number of people outreached as

$$x_{t,a} = \frac{(\delta_t I_{t-1} p_{t-1,U}) - (\mu I_{t-1} p_{t-1,U})}{\theta\varphi} \tag{32}$$

We calculate the unknown terms on the right-hand side of the above equation as follows,

$\mu = \frac{\text{number of persons newly diagnosed with HIV in 2010}}{\mathrm{I_{2010}p_{2010,U}}}$ by assuming that the rate of HIV diagnosis in the year 2010, which is just prior to the implementation of the first HIV national strategic plan , is the rate of diagnosis through voluntary testing and that it remains the same for future years,

$\theta = 30\%$ an assumption taken from , and

$$\varphi = \left(\frac{I_{t-1}}{N_{t-1}}\right)\left(\varphi_{InterventionGroup,year} / \frac{I_{Detroit,year}}{N_{Detroit,year}}\right),$$

where $\varphi_{InterventionGroup,year} = 0.0153$ is the proportion of persons testing positive in the outreach intervention program , and $\frac{I_{Detroit,year}}{N_{Detroit,year}} = 0.0057$ is the prevalence of HIV in Detroit at the time the study was conducted[1].

Estimation of $n_{t,a}$- Number of persons testing negative at time step t under action a:

Following from the previous subsection, we can write

$x_{t,a}\theta(1-\varphi)$ as the number of persons testing negative for HIV among those tested through an outreach intervention program, and

$\mu(N_t - I_{t-1}p_{t-1,U} - x_{t,a})$ as the number of persons testing negative for HIV among those who voluntarily get tested.

Thus, we estimate the total number of HIV negative cases as:

$$n_{t,a} = \mu(N_t - I_{t-1}p_{t-1,U} - x_{t,a}) + x_{t,a}\theta(1-\varphi) \quad (33)$$

The estimation of all parameters on the right side of the equation has been discussed earlier.

Estimation of X_v - Variable cost per person testing negative:

Let

$c_{r,n}$ be the rapid test cost for an HIV-negative case,

$c_{c,n}$ be the conventional test cost for an HIV-negative case,

n_n be the notification cost for an HIV-negative case,

c_{add} be the additional cost per person performed in a non-clinical setting,

τ be the proportion rapid test, and

α be the proportion tested in the clinical setting.

We can then estimate the variable cost per person for an HIV negative case as

$$X_v = \tau c_{r,n} + (1-\tau)c_{c,n} + n_n + (1-\alpha)c_{add} \quad (34)$$

For the parameters on the right-hand side of the equation, we use estimates from the literature and are summarized in Table A1.

Estimation of Y_v – Variable cost per person testing positive:

Let

$c_{r,p}$ be the rapid test cost for an HIV-positive case,

$c_{c,p}$ be the conventional test cost for an HIV-positive case,

$n_{r,p}$ be the rapid test notification cost for an HIV-positive case,

$n_{c,p}$ be the conventional test notification cost for an HIV-positive case, and

c_{cnf} be the confirmatory cost.

Hence, we estimate the variable cost per person for a HIV positive case as follows:

$$Y_v = \tau(c_{r,p} + n_{r,p}) + (1-\tau)(c_{c,p} + n_{c,p}) + c_{cnf} + (1-\alpha)\, c_{add} \quad (35)$$

For the parameters on the right-hand side of the equation, we use estimates from the literature and are summarized in Table A1.

Estimation of O_v- Variable cost per person outreached:

[1] This equation assumes that the outreach intervention program would be targeted towards higher-risk individuals and thus the proportion testing positive would be higher than the overall prevalence in the population, as evident by the data (presented above) in this study [5].

As the proportion of HIV-infected persons unaware of their infection decreases, the marginal cost of reaching one additional HIV-positive person increases. Therefore, we formulate the variable cost of outreach intervention as a non-linear function of increment in the proportion of people outreached relative to the number of people outreached in the base year with respect to the total population.

Let O_{v_0}be the base year variable cost for outreaching x_0 people among the total population in the base year (2015), x_0 be the number of people outreached in the base year estimated using data in with outreach intervention effectiveness of 30% , Δx be the increment in the number of populations outreached from the base year with respect to total population at time t, calculated as $\frac{x_t-x_0}{N_t}$, and w be the coefficient of a non-linear variable cost function.

Then per person variable cost of Δx increment in outreaching is:

$$O_v = O_{v_0} e^{\Delta x W} \tag{36}$$

Derivation:

We assume that Δx increase in the number of people outreached imposes an extra cost of $O_v w \Delta x$ and estimate per person outreach cost as,

$$O_{v+\Delta x} = O_v + O_v w \Delta x$$
$$O_{v+\Delta x} - O_v = O_v w \Delta x$$
$$\frac{O_{v+\Delta x} - O_v}{O_v} = w \Delta x$$

Integrating on both sides,

$$O_v = O_{v_0} e^{\Delta x W}. \tag{37}$$

For the parameters on the right-hand side of the equation, we use estimates from the literature and are summarized in Table A2.

Estimation of $X_{f-cl,a}$ – Total fixed cost of testing in a clinical setting corresponding to action a:

We assume that infected and uninfected persons share the same fixed costs in a clinical setting:

Let

α be the proportion of people testing in a clinical setting,

m_{cl} be capacity of a clinic (we use an average estimate for clinic capacity),

f_c be the total fixed cost per clinic with $\mathrm{m_{cl}}$ capacity, and

$X_{f-cl,a}$ be the total clinical fixed cost.

Then the total fixed cost for testing $\alpha(\mathrm{r}_{\mathrm{t},a} + \mathrm{n}_{\mathrm{t},a})$ number of people in a clinical setting is estimated as follows:

$$X_{f-cl,a} = \left\lceil \frac{(r_{t,a} + n_{t,a})\,\alpha}{m_{cl}} \right\rceil f_c \tag{38}$$

For the parameters on the right-hand side of the equation, which were not discussed earlier, we use estimates from the literature , and are summarized in Table A1.

Table A1. Cost Components of HIV Positive and Negative Testing[2].

Input	Value	Source
Testing cost, including screen and confirmatory test		
HIV-uninfected individual, rapid screen	\$22.13 (2012\$)	Based on cost components from .
HIV-uninfected individual, conventional screen	Before 2016: \$8.24 (2012\$) 2016+: \$10.36 (2012\$)	Based on cost components from .
HIV-infected individual, rapid screen	Before 2016: \$86.70 (2012\$) 2016+: \$78.80 (2012\$)	Based on cost components from and adjusted to 2012\$. Assumes a repeat screen and a Western blot confirmatory test before 2016; a Geenius HIV 1/2 confirmatory screen is used in 2016 and after.
HIV-infected individual, conventional screen	Before 2016: \$60.02 (2012\$) 2016+: \$58.91 (2012\$)	Based on cost components from and adjusted to 2012\$. Assumed Western blot confirmatory test before 2016, and a Geenius HIV 1/2 confirmatory screen in 2016 and after.
NAT[3], applied for discrepant Western blot confirmatory test	\$160.07 (2012\$)	
Additional cost per test performed in non-clinical (vs. clinical) setting	\$52.66 (2005\$)	
Notification costs		
HIV-uninfected	\$0.45 (2009\$)	
HIV-infected, conventional screen	\$5.88 (2009\$)	
HIV-infected, rapid screen	\$10.86 (2009\$)	
Outreach cost per test (when applied)	\$13.67 (2005\$)	

Estimation of $X_{f-ncl,a}$ – Total fixed cost of testing in a non-clinical setting corresponding to action a:

[2] The component of testing costs are taken from HOPE model technical report [6].

3 Note: NAT = HIV nucleic acid amplification test

We assume that infected and uninfected persons share the same fixed costs in non-clinical testing.

Let m_{nc} be the average capacity of a non-clinical setting and f_{nc} be the total fixed cost per non-clinic setting with capacity m_{nc}.

Then total fixed cost to test $(1-\alpha)(r_{t,a}+n_{t,a})$ people in a non-clinical setting is estimated as follow:

$$X_{f-ncl,a}=\left\lceil\frac{(r_{t,a}+n_{t,a})\,(1-\alpha)}{m_{nc}}\right\rceil f_{nc} \tag{39}$$

For the parameters on the right-hand side of the equation that were not discussed earlier, we use estimates from the literature , and are summarized in Table A1.

Estimation of $X_{f-o,a}$- Fixed cost of outreach intervention:

Let

m_o be the capacity of outreach intervention, and

f_o be the total fixed cost of outreaching x_t people.

Then the total fixed cost to outreach $x_{t,a}$ number of people is estimated as

$$X_{f-o,a}=\left\lceil\frac{x_t}{m_o}\right\rceil f_o \tag{40}$$

For the parameters on the right-hand side of the equation that were not discussed earlier, we use estimates from the literature and are summarized in Table A2.

Table A2. Summary of Parameters and Values Used in Testing and Outreach Intervention.

Notation	Description	Value(s)	Reference
O_{v_0}	Base year variable cost for outreaching x_0 people	\$16.59 (2015\$)	
Δx	Increment in the number of populations outreached from the base year		Calculated in PATH
W	coefficient of variable cost	0.1, 0.2, 0.3	Assumption
m_{cl}, m_{nc}	Clinical/ non-clinical capacity	1000	Assumption
f_c	Total fixed cost per clinic with m_{cl} capacity	\$56379 (2015\$)	
f_{nc}	Total fixed cost per clinic with m_{cl} capacity	\$64851 (2015\$)	

Estimation of retention-in-care cost

We assume that the required proportion on ART under any action (a) can be achieved by implementing retention-in-care programs that ensure patients remain in care and consistently take antiretroviral therapy treatment to achieve viral load suppression . Retention-in-care programs could include a combination of a face-to-face meeting with patients at primary care visits, brief interim phone contacts between appointments, appointment reminders, and missed-visit calls .

We assume retention-in-care programs could include fixed and variable costs. Fixed costs include the cost of office space, durable items such as computers, printers, telephones, etc. And variable costs include staff time spent on management of patients and personal contact with patients. We assume the variable cost per person is a non-linear function of the proportion of people retained in care, i.e., the marginal cost to achieve one additional person retained in care increases as the proportion on ART increases indicating more efforts would be needed to retain a larger number of people in care.

We estimate the cost of retention-in-care corresponding to an action as follows.

Let

$d_{t,a}$ be the number of persons retained in care at time step t under action a;

R_v be the variable cost per person retained in care,

$E_{f,a}$ be the total fixed cost of retaining in care under action a, and

C_{a_r} be the total cost of retention in care under action a.

Then, we can calculate the total cost retention in care action as:

$$C_{a_r} = d_{t,a}R_v + E_{f,a}. \tag{41}$$

Below we discuss the estimation of each component on the right-hand side of the equation.

Estimation of $\boldsymbol{d_{t,a}}$- Number of people retained in care:

Note that, this is the number of persons successfully intervened (number retained-in-care) and estimated in Appendix Section 2, i.e.,

$$d_{t,a} = (1 - \rho_t)I_{t-1}p_{t-1,ART} \tag{42}$$

The parameters on the right-hand side of the equation are calculated in the PATH 2.0 simulation model.

Estimation of $\boldsymbol{R_v}$- Retention-in-care variable costs:

Let

p_{ART_0} be the proportion on ART in the base year 2015, which is calculated in PATH,

R_{ART_0}be the base year variable cost of achieving $\mathrm{p_{ART_0}}$,

Δp_{ART} is the increment in the proportion on ART with respect to the base year, and

Y is the coefficient of the non-linear variable cost function.

Similar to testing outreach variable cost, we calculate the retention-in-care variable cost as follows:

$$R_{ART} = R_{ART_0}e^{\Delta p_{ART}Y} \tag{43}$$

For the parameters on the right hand-side of the equation that were not discussed earlier, we use estimates from the literature and are summarized in Table A3.

Estimation of $\boldsymbol{E_{f,a}}$-Retention-in-care fixed costs:

Let

m_r be the retention-in-care intervention setting capacity, and

f_r be the fixed cost per retention-in-care intervention program (with capacity m_r).

We estimate the total fixed cost for retaining $d_{t,a}$ number of people in care as:

$$E_{f,a} = \left\lceil \frac{d_{t,a}}{m_r} \right\rceil f_r \tag{44}$$

For the parameters on the right hand-side of the equation that were not discussed earlier, we use estimates from the literature and are summarized in Table A3.

Table A3. Summary of Parameters Used in Retention-in-care Intervention.

Notation	Description	Value(s)	Reference
p_{ART_0}	Proportion on ART at time step 0		Calculated in PATH
R_{ART_0}	Base year variable cost of achieving p_{ART_0}	\$117 (2015\$) \$235 (2015\$) \$300 (2015\$)	Low, medium, and high retention-in-care costs
Δp_{ART}	Increment in the proportion on ART		Calculated in PATH
Y	Coefficient of variable cost	0.1, 0.2, 0.3	Assumption
m_r	Retention-in-care intervention program capacity	500	Assumption based on
f_r	Fixed cost per retention-in-care intervention program with capacity m_r	\$17977 (2015\$) \$22708 (2015\$) \$29330 (2015\$)	Low, medium, and high retention-in-care costs

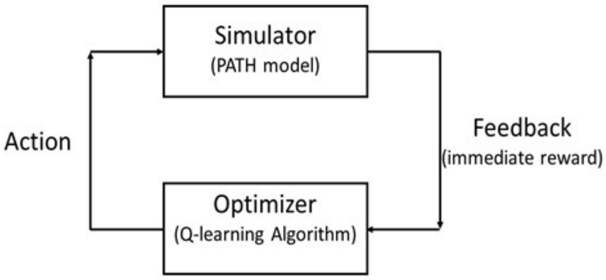


Figure A3. Schematic of the reinforcement learning (RL) algorithm. The RL algorithm takes action, and it is fed into the simulator. The simulator simulates the action and returns an immediate reward to the algorithm to update the following action accordingly.

5. Q-Learning algorithm

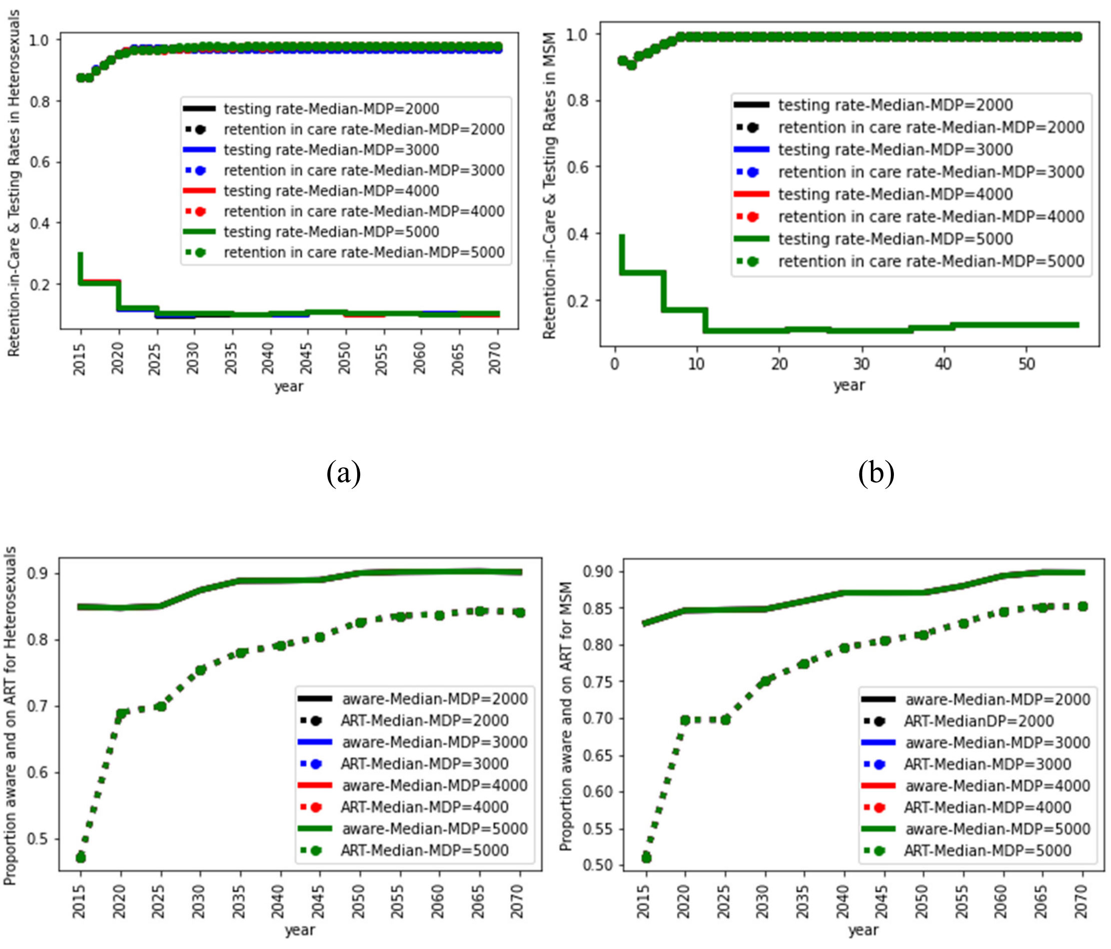
Table A4. Finite-horizon Q-learning Algorithm to Identify Optimal Solutions to the Markov Decision Process .

```
• Set $k\ = 0, T = 11$ (maximum number of decision stages)
• Initialize Q-factor, $Q_k\ (x, t, a)$ to a very small negative number (i.e.$-10^{-15}$)
• While $k < k_{max}$
1. Set $t = 1$ and current state as $x$, setting PATH to start at HIV epidemic in the year 2015.
2. Select action, using decaying epsilon greedy ,
       $a(x,t) = \begin{cases} random\ selection\ from\ action\ space\ A\ with\ probability\ \epsilon_t \\ arg \max_{b \in A(x,t)} Q_k(x,t,b)\ with\ probability\ (1-\epsilon_t) \end{cases}$
       $\epsilon_t = \begin{cases} \frac{0.85}{k} + 0.049 & k \leq 4000 \\ \frac{0.85}{k-3500} + 0.049 & k > 4000 \end{cases}$
3. Simulate the system using PATH, result in the transition state in $t+1$ to be $x'$.
4. Using the PATH model, calculate the immediate reward, $R_a(x, x')$.
5. Update $Q(x, t, a)$ as
       $Q_{k+1}(x,t,a) = (1-\alpha)Q_k(x,t,a) + \ \alpha \left[R_a(x,x') + \gamma \max_{b \in A(x',t+1)} Q_k(x',t+1,b)\right]$
   where,
       ➢ $\alpha$ is the step size a polynomial rate decaying with $k$ . We assumed
         $\alpha = \frac{1}{(k+1)^{0.5625}}$
       ➢ $\gamma$ is the discount factor (We assume $\gamma$ =1).
6. Set $x \leftarrow x'$, and $t = t + 1$
       ➢ if $t = T$ set $k = k + 1$, and go to step 1
       ➢ else, go to step 2
• Identify optimal solution: optimal action "$a$" to be taken at time "$t$" when the system is in
  state $x$:
       $a(x,t) \epsilon\ arg \max_{b \in A(x,t)} Q_{k_{max}}\ (x,t,b)$.
```

6. Testing convergence of the Q-learning reinforcement algorithm

In this section, we present the uncertainty in results, specifically, the uncertainty in the optimal policy $\left[a_{unaware,HET}, a_{ART,HET}, a_{unaware,MSM}, a_{ART,MSM}\right]$ and the uncertainty in values of testing and retention-in-care rates $[\delta, 1-\rho]$, corresponding to each scenario for different values of Q-learning training iterations k_{max} (2000, 3000, 4000, and 5000). An algorithm is said to have converged if, it has reached a local optima through the iterative search process, i.e., successfully solved for an optimal combination of testing and retention-in-care rates. If the number of iterations is not sufficiently large, there is a risk that the algorithm is terminated before convergence. The ideal number is typically determined through experimentation.

Further, there could be multiple local optima, i.e., multiple policies could yield similar total rewards, and because of the stochastic nature of the epidemic system, the optimal policy could be a range rather than a point estimate. Therefore, we ran the model for varying number of iterations, 2000, 3000, 4000, and 5000, and compared the corresponding total rewards, to ensure convergence and obtain the uncertainty range in optimal policies (Figure A4, A7, and A10 for each of the three cost functions). Note also that the training with $k_{max} = 5000$ has an additional exploration after 4000 iterations (defined by the epsilon-greedy action selection defined in the main manuscript). The relative difference in the total population costs between the varying iterations were at most 2% in each cost function evaluated, suggesting convergence (Figures A6, A9, and A12 for each of the three cost functions). The changes in new infections and the number of people with HIV (PWH) were also minimal (Figures A5, A8, and A11 for each of the three cost functions). The corresponding optimal policies differed slightly, more so in future years than earlier years, suggesting stochastic uncertainty as the model projects further into the future (Figure A4, A7 and A10 for each of the three cost functions). Therefore, in Results of the main paper, we present the range of optimal policies across these iterations as the uncertainty range.



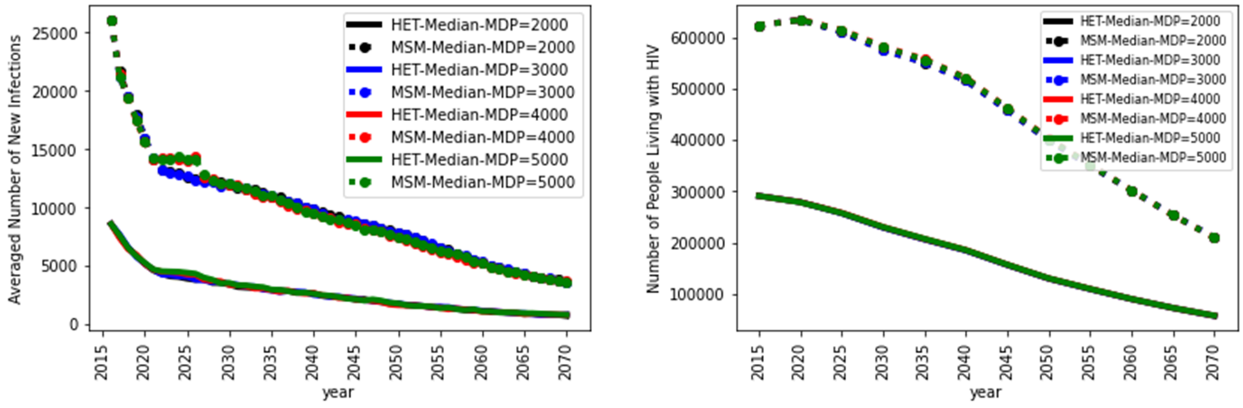
(a)

(b)

(c)

(d)

Figure A4. Medium Cost Function: Combination of optimal policy (testing and retention-in-care rates) for heterosexuals (a) and MSM (b), and corresponding proportion of aware and on ART for heterosexuals (c) and MSM (d) for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green). Results are an average of 100 runs.



(a) (b)

Figure A5. Median Cost Function: Impact of implementing a combination of optimal policy on the number of new infections (a) and the number of people living with HIV (b) for heterosexuals (solid lines) and MSM (dashed lines) for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green). Results are an average of 100 runs.

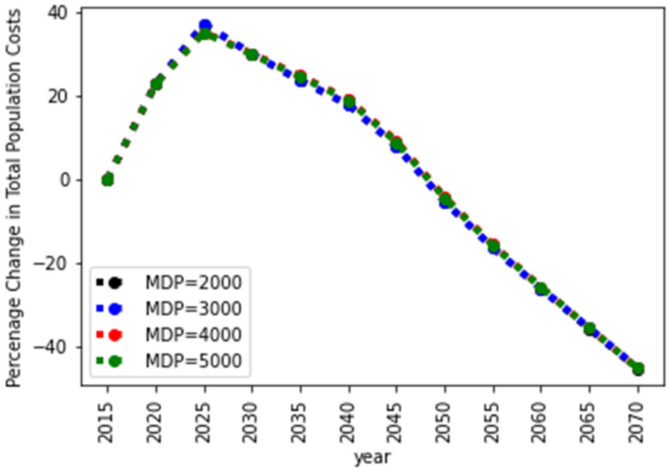
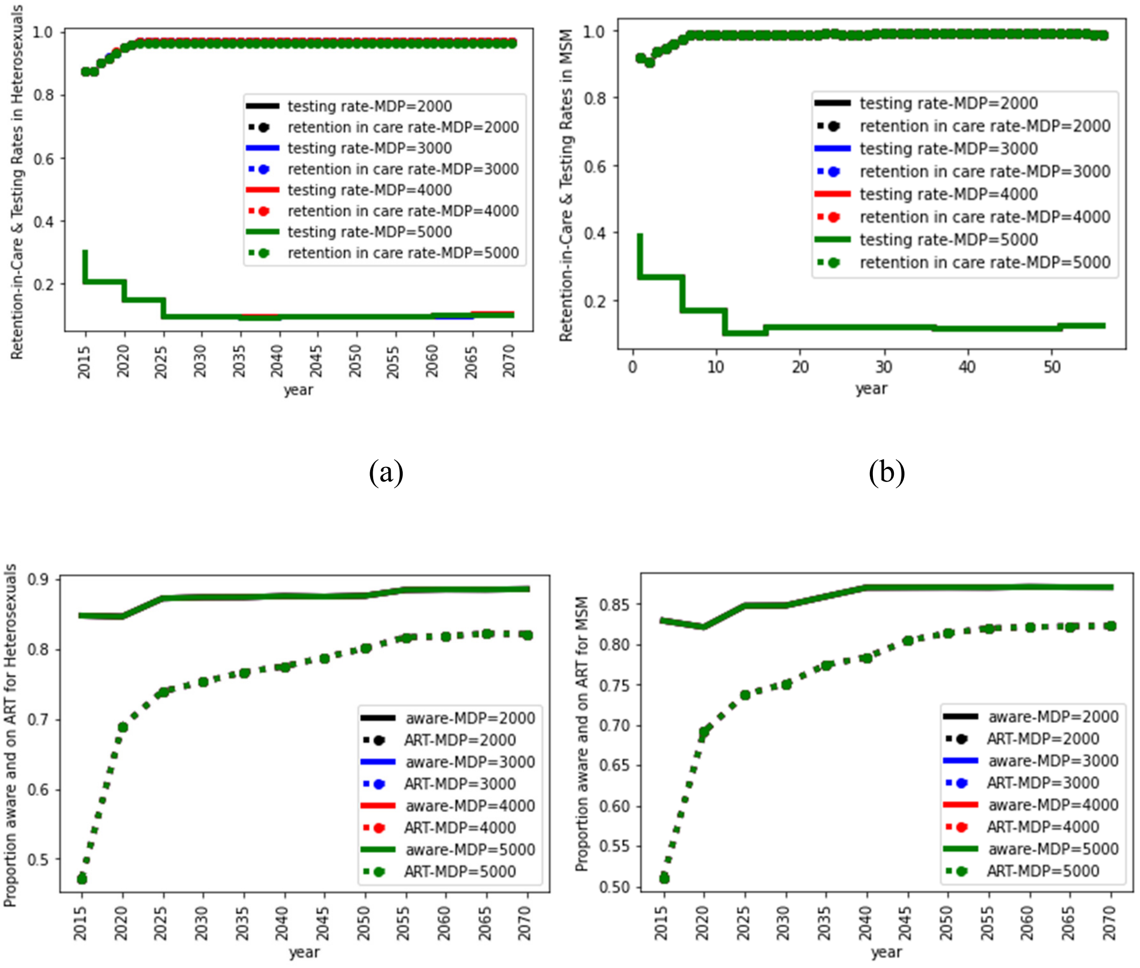


Figure A6. Median Cost Function: Percentage increment in total population cost of implementing optimal policy for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green). Results are an average of 100 runs.



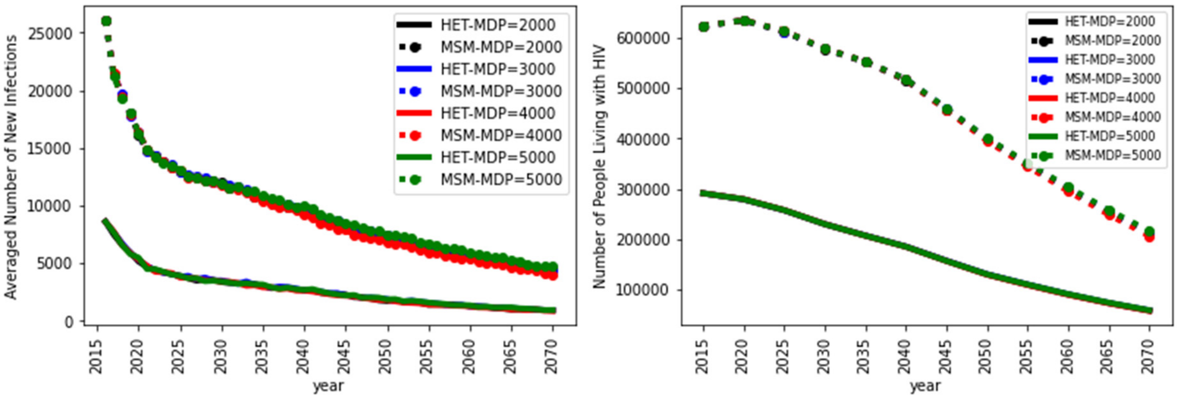
(a)

(b)

(c)

(d)

Figure A7. HTLR Cost Function: Combination of optimal policy (testing and retention-in-care rates) for heterosexuals (a) and MSM (b), and corresponding proportion of aware and on ART for heterosexuals (c) and MSM (d) for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green). Results are an average of 100 runs.



(a) (b)

Figure A8. HTLR Cost Function: Impact of implementing combination of optimal policy on the number of new infections (a) and the number of people living with HIV (b) for heterosexuals (solid lines) and MSM (dashed lines) for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green).

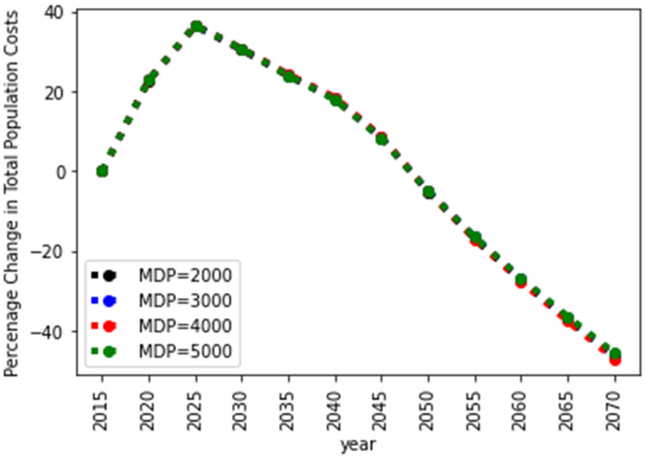
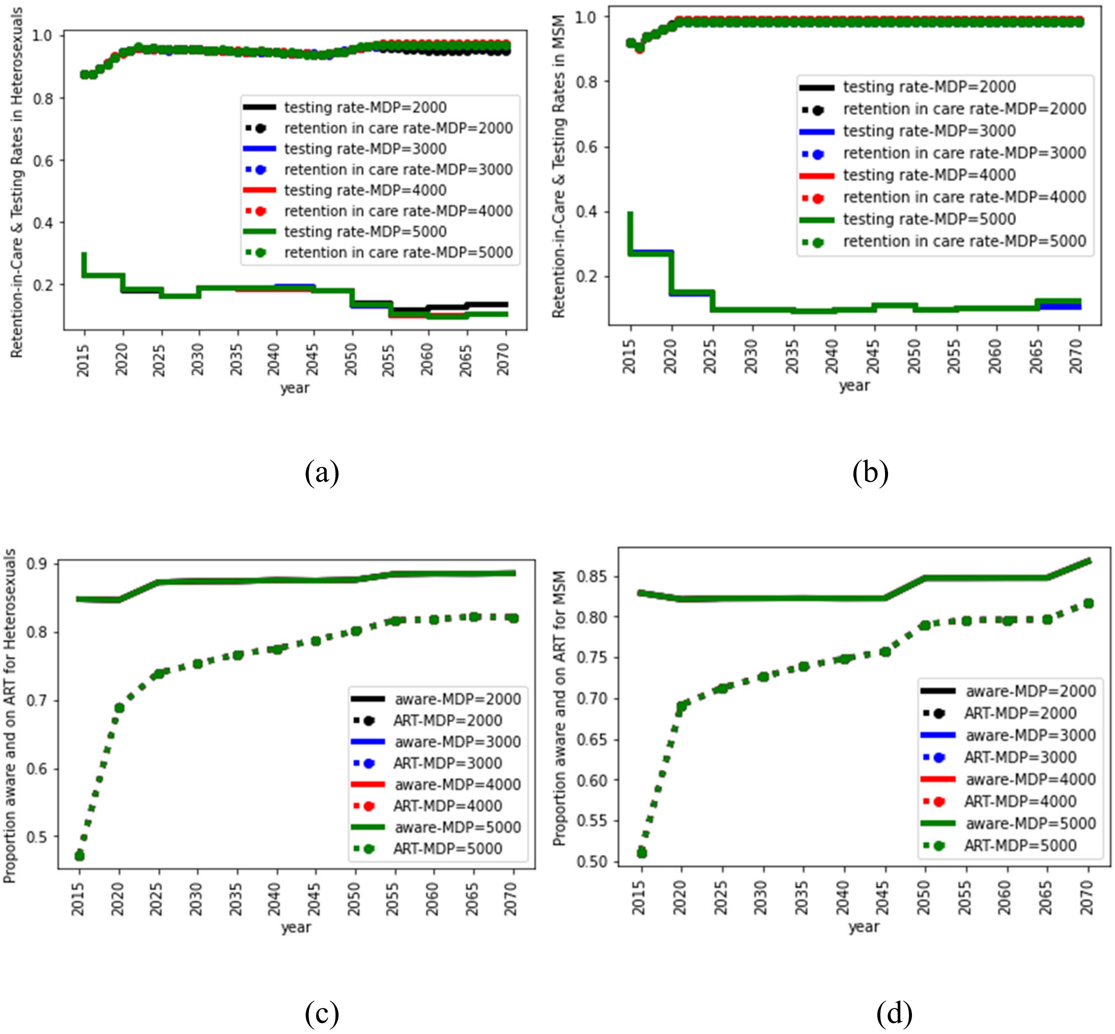


Figure A9. HTLR Cost Function: Percentage increment in total population cost of implementing optimal policy for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green). Results are an average of 100 runs. Results are an average of 100 runs.



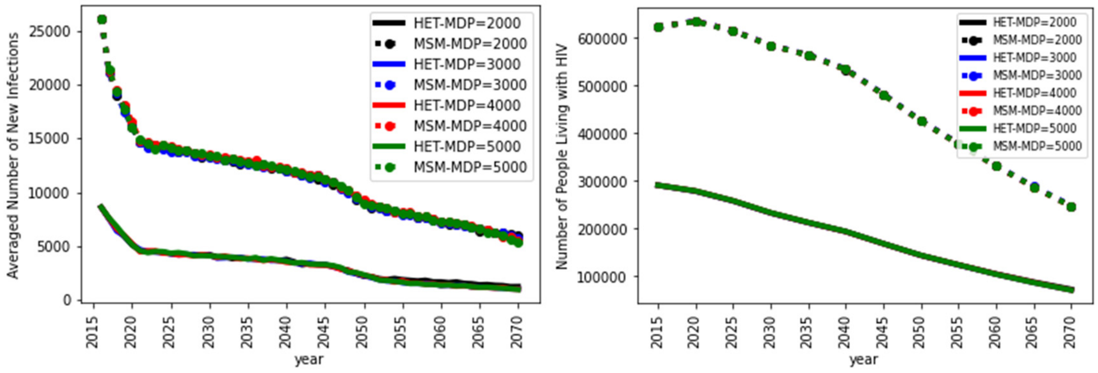
(a)

(b)

(c)

(d)

Figure A10. LTHR Cost Function: Combination of optimal policy (testing and retention-in-care) for heterosexuals (a) and MSM (b), and corresponding proportion of aware and on ART for heterosexuals (c) and MSM (d) for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green). Results are an average of 100 runs.



(a) (b)

Figure A11. LTHR Cost Function: Impact of implementing combination of optimal policy on the number of new infections (a) and the number of people living with HIV (b) for heterosexuals (solid lines) and MSM (dashed lines) for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green). Results are an average of 100 runs.

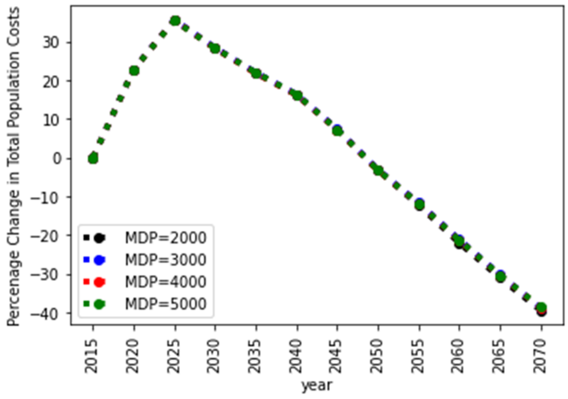
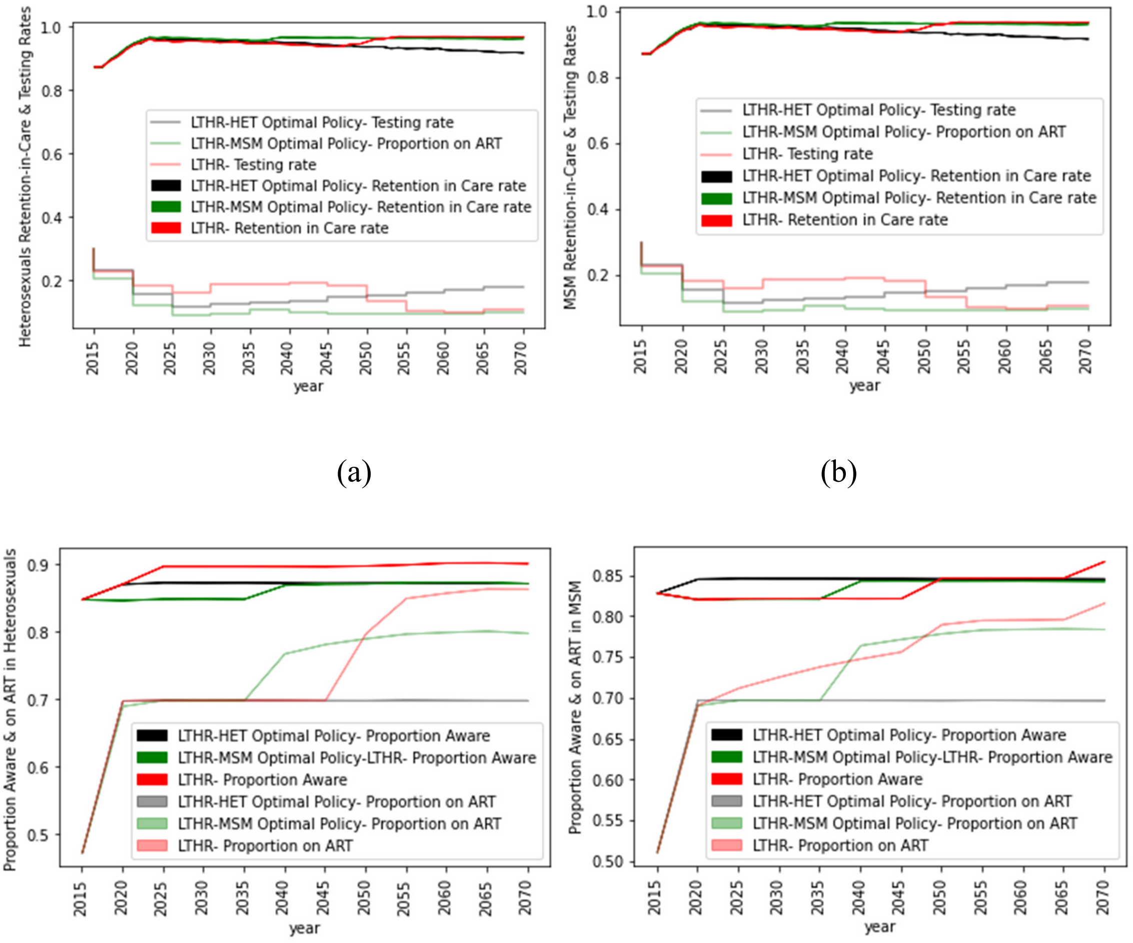


Figure A12. LTHR Cost Function: Percentage increment in total population cost of implementing optimal policy for MDP iterations of 2000 (black), 3000 (blue), 4000 (red), and 5000 (green). Results are an average of 100 runs.

7. Evaluating the optimality of the counter-intuitive results under the LTHR MSM scenario

As discussed in the manuscript, comparing across the cost function assumptions, the optimal rates were generally intuitive for heterosexuals, highest testing and lowest retention-in-care in LTHR (which assumed lowest unit cost for testing and highest unit cost for retention-in-care) and lowest testing and highest retention-in-care rates in HTLR, though the differences in retention-in-care rates were modest (paper Figure 1a). However, for MSM, though the model suggested optimal rates were similar in all three cost functions, it counter-intuitively suggested a slightly lower testing rate in LTHR compared to Median and HTLR. It suggested to instead spend those resources on maintaining retention-in-care rates at the level of Median and HTLR (paper Figure 1b), such that the proportion of MSM on ART in LTHR, though lower than in Median and HTLR, was higher than the proportion of heterosexuals on ART in LTHR (paper Figure 1d). The optimality of this counter-intuitive strategy in MSM was evaluated by a counterfactual simulation run using the optimal LTHR strategy of heterosexuals for both heterosexuals and MSM (Figure A13). The number of new infections, PWH, and costs were higher in the counterfactual simulation, confirming the optimality of the policy (Figures A14 and A15). The results of this counterfactual run also suggest that the reasoning behind the counter-intuitive strategy in MSM is likely because of the higher prevalence of HIV in MSM, i.e., lowering the spending on retention-in-care would lead to more infections than lowering the spending on testing. Note also in this counterfactual run, the results for heterosexuals also change over time compared to the original (though they both start with the same optimal strategy) because of the dynamics of contact mixing between MSM and heterosexuals over time and the functionality of the model to correct for those dynamic changes and find a new optimal. A similar counterfactual run of using the MSM optimal strategy on the heterosexual risk group generated similar findings.



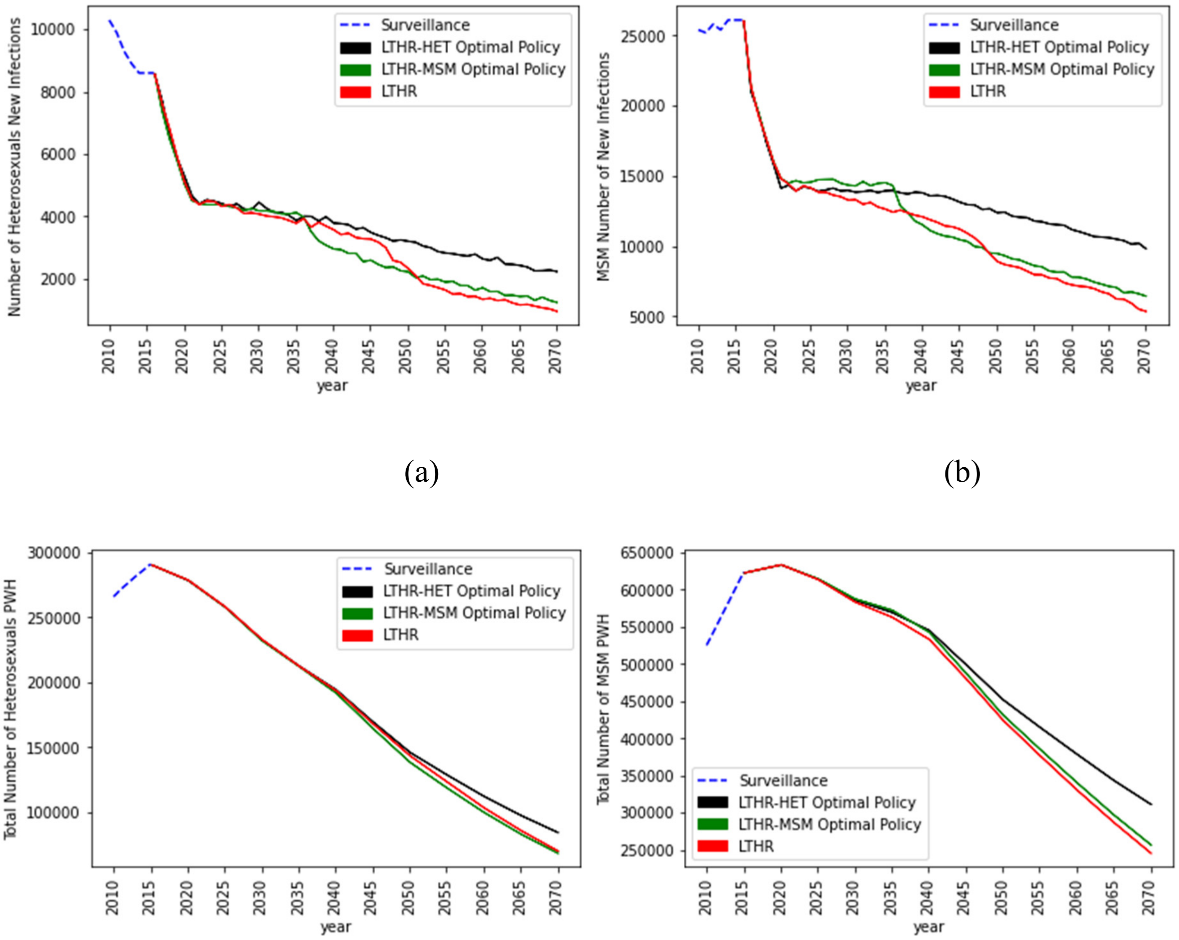
(a)

(b)

(c)

(d)

Figure A13. Top: Optimal combination of testing rate (a) and retention-in-care rate (b). Bottom: Corresponding proportion aware (c) and proportion on ART (d).



(a) (b)

(c) (d)

Figure A14. Top: Number of new infections for heterosexuals (a) and MSM (b). Bottom: Number of people living with HIV for heterosexuals (a) and MSM (b).

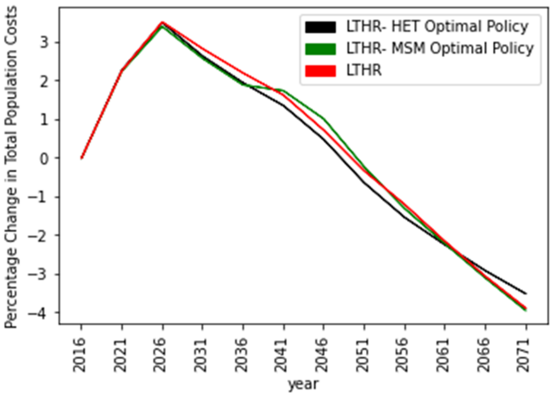


Figure A15. Comparing percentage change in total population cost of LTHR cost function (red) with heterosexual LTHR cost function applied to both risk groups (black) and with MSM LTHR cost function applied to both risk groups (green) all with MDP = 5000.

Reference

1. A. Gosavi, Simulation-based optimization: Parametric optimization techniques and reinforcement learning, 2003.
2. C. Gopalappa, P. G. Farnham, Y.-H. Chen, S. L. Sansom, Progression and Transmission of HIV/AIDS (PATH 2.0) A New, Agent-Based Model to Estimate HIV Transmissions in the United States, *Med. Decis. Making.*, **37** (2017), 224–233.
3. *Centers for Disease Control and Prevention*, Monitoring selected national HIV prevention and care objectives by using HIV surveillance data-United States and 6 US dependent areas-2010, *Citeseer*, 2012.
4. E. M. Gardner, M. P. McLees, J. F. Steiner, C. del Rio, W. J. Burman, The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection, *Clini. Infect. Dis.*, **52** (2011), 793–800.
5. R. K. Shrestha, H. A. Clark, S. L. Sansom, B. Song, H. Buckendahl, C. B. Calhoun, et al., Cost-effectiveness of finding new HIV diagnoses using rapid HIV testing in community-based organizations, *Public Health Rep.*, **123** (2008), 94–100.
6. E. U. Jacobson, K. A. Hicks, E. L. Tucker, P. G. Farnham, S. L. Sansom, Effects of reaching national goals on HIV incidence, by race and ethnicity, in the United States, *J. Public Health Manag. Pract.*, **24** (2018), E1–E8.
7. L. K. W. D. Agero, R. D. Gorsky, G. M. Seeman, Cost of outreach for HIV prevention among drug users and youth at risk, *Drugs Soc.*, **9** (1996), 185–197.
8. L. Guinness, L. Kumaranayake, K. Hanson, A cost function for HIV prevention services: is there a'u'--shape?. **5** (2007), 13.
9. L. Kumaranayake, The economics of scaling up: cost estimation for HIV/AIDS interventions, *Aids*, **22** (2008), S23–S33.
10. S. D. Pinkerton, L. M. Bogart, D. Howerton, S. Snyder, K. Becker, S. M. Asch, Cost of rapid HIV testing at 45 US hospitals, *AIDS Patient Care STDs*, (2010), 409–413.
11. P. G. Farnham, A. B. Hutchinson, S. L. Sansom, B. M. Branson, Comparing the costs of HIV screening strategies and technologies in health-care settings, *Public Health Rep.*, **123** (2008), 51–62.
12. A. Lansky, J. Christopher, O. Emeka, S. Catlainn, M. P. Joyce, E. DiNenno, et al., Estimating the Number of Heterosexual Persons in the United States to Calculate National Rates of HIV Infection, *PloS One*, **10** (2015), e0133543.
13. A. Chandra, V. G. Billioux, C. Copen, C. Sionean, HIV risk-related behaviors in the United States household population aged 15-44 years: data from the National Survey of Family Growth, 2002 and 2006-2010, *Natl. Health Stat. Rep.*, **46** (2012), 1–19.
14. N. Khurana, E. Yaylali, P. G. Farnham, K. A. Hicks, B. T. Allaire, E. Jacobson, et al., Impact of Improved HIV Care and Treatment on PrEP Effectiveness in the United States, 2016-2020, *J. Acquir. Immune Defic. Syndr*, **78** (2018), 399–405.
15. *White House*, National HIV/AIDS strategy for the United States 2010. Available from: https://www.hivlawandpolicy.org/sites/default/files/NHAS%20strategy.pdf. Accessed April 2020.

16. A. B. Hutchinson, P. G. Farnham, S. B. Lyss, D. A. White, S. L. Sansom and B. M. Branson, Emergency department HIV screening with rapid tests: a cost comparison of alternative models, *AIDS Educ. Prev.,* **23** (2011), 58–69.
17. A. B. Hutchinson, S. F. Ethridge, L. G. Wesolowski, R. K. Shrestha, M. Pentella, B. Bennett, et al., Costs and outcomes of laboratory diagnostic algorithms for the detection of HIV, *J. Clin. Virol.,* **58** (2013), e2–e7.
18. R. K. Shrestha, L. Gardner, G. Marks, J. Craw, M. Faye, T. P. Giordano, et al., Estimating the cost of increasing retention in care for HIV-infected patients: results of the CDC/HRSA retention in care trial, *J. Acquir. Immune Defic. Syndr.*, **68** (2015), 345.
19. D. Kumari, M. Chaudhary, A. K. Mishra. An Optimal Transfer of Knowledge in Reinforcement Learning through Greedy Approach. 2019.
20. T. Nieuwdorp, *Dare to Discover: The Effect of the Exploration Strategy on an Agent's Performance,* (2017).
21. E. Even-Dar, Y. Mansour, Learning rates for Q-learning, *J. Mach. Learn. Res.*, **5** (2003), 1–25.