Submitted to *Manufacturing & Service Operations Management*
manuscript MSOM

The Price of Nonabandonment: HIV in Resource-Limited Settings

Amin Khademi
Clemson University, e-mail: khademi@clemson.edu

Denis R. Saure
University of Chile, e-mail: dsaure@dii.uchile.cl

Andrew J. Schaefer
University of Pittsburgh, e-mail: schaefer@pitt.edu.

Ronald S. Braithwaite
New York University, e-mail: Scott.Braithwaite@nyumc.org.

Mark S. Roberts
University of Pittsburgh, e-mail: mroberts@pitt.edu.

The global fight against HIV/AIDS is hindered by a lack of drugs in the developing world. When patients in these countries initiate treatment, they typically remain on it until death, thus policy makers and physicians follow *nonabandonment* policies. However, treated patients develop resistance to treatment, so in many cases untreated patients might benefit more from the drugs. In this paper we quantify the opportunity cost associated with restricting attention to nonabandonment policies. For this, we use an approximate dynamic programming framework to bound the benefit from allowing premature treatment termination. Our results indicate that in Sub-Saharan Africa, the price associated with restricting attention to nonabandonment policies lies between 4.4% and 8.1% of the total treatment benefit. We also derive superior treatment allocation policies, which shed light on the role behavior and health progression play in prioritizing treatment initiation and termination.

Key words: HIV, resistance, nonabandonment allocation policy, approximate dynamic programming.

1. Introduction

Motivation. The human immunodeficiency virus (HIV) causes acquired immunodeficiency syndrome (AIDS), a condition that progressively reduces the effectiveness of the immune system. There are currently 35.3 million people infected with HIV worldwide, and HIV/AIDS has killed over 36 million since 1981. In 2012, 2.3 million became infected and 1.6 million died of AIDS-related causes (AIDS.gov, 2013). Antiretroviral therapy (ART), the only treatment option for chronic HIV infection, temporarily impedes the progression of the virus. Unfortunately, the longer a patient receives treatment, the more likely she is to develop resistance to drugs, thus reducing their effectiveness (PLATO Collaboration, 2004). Resistance occurs as the virus mutates in reaction to antiretroviral

drugs. As genotype testing for HIV drug resistance is typically unavailable in most resource-limited settings (Revell et al., 2013), mutation is not observable in practice and can only be inferred.

In the developed world, ART is available to those who require it (WHO, 2010b), and physicians follow *nonabandonment* treatment practices: Patients are typically treated until death irrespective of resistance, and treatment is discontinued only for palliative purposes (WHO, 2010a). This situation stands in stark contrast to that in the developing world, where access to ART is limited, and demand for treatment largely exceeds its supply. In Sub-Saharan Africa, where more than 22 million are currently infected, coverage (the number of patients treated over the total number of patients in need of treatment) for most countries is below 40% (Joint United Nations Programme on HIV/AIDS, 2010), and millions in need of treatment do not receive it (Joint United Nations Programme on HIV/AIDS, 2010). Despite this contrast, like in the developed world, patients in the developing world are typically treated until death (WHO, 2012), even after developing resistance, ignoring the fact that many untreated patients might benefit more from treatment.

Main objectives. This work aims to shed light on the cost (in terms of population life-years) of continuing treatment of patients who receive only marginal benefit from therapy over reallocating it to those whom would derive significantly more benefit. That is, we aim to evaluate the opportunity cost associated with nonabandonment policies, which we refer to as the price of nonabandonment (PoN). In the process, we analyze properties of efficient treatment allocation policies, for both nonabandonment and abandonment-permitting scenarios.

To achieve our objectives we develop a mathematical framework to quantify PoN in settings where access to ART is limited. In particular, we use an approximate dynamic programming (ADP) approach to develop efficient treatment allocation policies when both nonabandonment policies and abandonment-permitting policies are considered. The exact computation of PoN is intractable (our Markov decision process (MDP) formulations consider unbounded state spaces). However, the ADP framework allows us to estimate a lower bound on PoN by computing the difference between the performance of arbitrary abandonment-permitting policies (we consider those emanating from our ADP approximation), and an upper bound on the performance of an optimal policy in the nonabandonment setting. We develop one such bound by relaxing some of the effects of HIV and ART in transmission and health progression, and another bound from our ADP approximation.

Ethical Considerations. Departing from nonabandonment by possibly withholding therapy involves potentially difficult ethical questions, as one may wonder whether it is morally permissible to remove patients from care so that others may benefit more. In this regard, this paper does not attempt to answer those questions but rather to characterize some of the many trade-offs present when trying to answer said questions.

Medical ethicists have explored various mechanisms for allocating scarce therapies. Using the taxonomy proposed by Persad et al. (2010), PoN can be viewed as the value of moving from prioritarianism (which prioritizes the sickest) to utilitarianism (which maximizes the societal benefits of the scarce therapies), akin to recent changes in prioritization for kidney transplantation, from a *sickest-first* rule to one that takes post-transplant survival into account (Caplan, 1995; Organ Procurement and Transplantation Network, 2014); see Section 2. In particular, we consider a utilitarian approach that permits triage decisions, including withdrawing treatment from individuals as needed. Although viewing ART allocation in resource-limited environments from a triage point of view is novel, and a contribution of the paper, the salient dimensions of such an approach have been firmly established in the medical ethics community. Using arguments grounded in the medical ethics literature, we provide more details of the ethical ramifications of moving from prioritarianism to utilitarianism in Section 2.

Main contributions and results. The main contribution of this paper is the development of a flexible framework for estimating PoN in resource-limited settings: We estimate that in settings with treatment coverage around 45% (recall that coverage is defined as the number of patients treated over the total number of patients in need of treatment), PoN lies between 4.4% and 8.1% of the total "benefit" from drugs. In particular, for current prevalence estimates, abandonment-permitting policies perform at least 5% better than nonabandonment policies when performance is measured in total discounted population life-years, relative to those achieved by following the guidelines proposed by the World Health Organization (WHO): In settings with treatment coverage around 45%, the ADP policies in abandonment-permitting and nonabandonment settings outperform WHO guidelines by 8% and 3%, respectively.

From a methodological standpoint, our work develops an ADP-based framework to bound PoN, and in doing so, we develop an algorithmic procedure for generating treatment allocation policies. The core of the procedure lies in approximating the value function associated with an MDP formulation of the treatment allocation problem using an affine approximating architecture. We show that when restricted to such an approximating architecture, "optimal" policies are of the *state-dependent* priority type, and this latter fact allows us to combine the linear programming (LP) approach to ADP with constraint sampling and decomposition techniques to iteratively approximate the solution to the MDP formulation (see Section 4.1). In addition, we develop a novel upper bound on the performance of optimal allocation policies by considering a relaxation that essentially separates the effects of treatment allocation on HIV transmission and patients' health progression: The bound follows from solving two simpler allocation problems (see the details in Section 4.2). This allows us to bound PoN, and also to assess the efficiency of the ADP-based policies: Under our

modeling assumptions, when coverage is close to that observed in Sub-Saharan Africa, our policies significantly outperform the WHO guidelines in abandonment-permitting settings (see Table 3 for details). In addition, we show that such results are robust to many assumptions made in the model.

To further assess the validity of our results, we compare the performance of the ADP-based policies against that of relevant benchmark under much *milder* assumptions than those made by the MDP model. For this we develop a large-scale simulation model that extends a clinically validated individual progression model and incorporates viral transmission. Through this model, we test policy performance in settings much closer to reality, thus further highlighting the high cost of nonabandonment practice. Our results here suggest that PoN ranges between 4.4% to 8.1% in settings representative of current conditions in Sub-Saharan Africa, depending on HIV prevalence. This estimate range, however, depends in the future availability of treatment funds.

Our results suggest that the advantage of abandonment-permitting policies emanates mainly from their ability to reallocate therapy away from the sickest patients toward healthier and risky patients. In this regard, our results provide insight on how health and behavior of patients should influence treatment initiation and termination. For example, for the case of nonabandonment policies, our results indicate that treatment initiation of sickest patients should be prioritized only when coverage and prevalence are relatively high; otherwise - all other things equal - priority should be given to patients with moderate health (this stands in contrast to WHO guidelines that call for *always* prioritizing the sickest patient). We summarize our findings in this regard in Section 7.2.

More broadly, our results should inform decision making in other settings where demand for treatment exceeds the supply as is often seen, for example, in the case of infectious diseases, especially in epidemics and pandemics, where resources, even in rich countries, become relatively scarce (Kirby, 2010; Rottman et al., 2010; Tabery and Mackett, 2008). There, the proposed framework should allow to develop policies to assign certain critical care resources (e.g., ventilator services).

2. Literature Review

HIV treatment has attracted the attention of many researchers, with early work focusing on optimizing treatment initiation: See Phillips et al. (2003) and references therein. Badri et al. (2006) consider a Markovian model for estimating the life expectancy and treatment cost of several treatment initiation policies, and Shechter et al. (2008) model therapy initiation as an optimal stopping problem. Wein et al. (1997) present a control-theoretic model to study single-patient optimal ART sequencing, assuming viral load (VL, a measure of the amount of virus in a mL of blood) is observable, and multiple ART regimes (a regimen typically combines three types of antiretroviral drugs affecting virus replication in different ways). Khademi et al. (2014) investigate the impact of new ART development on the optimal timing of HIV treatment. These studies consider a single patient

and assume ART is always available. In our study, drug availability and treatment initiation are, to some extent, endogenously determined.

Epidemic modeling has been widely studied in the clinical literature, and many models of transmission dynamics (key to studying the propagation of infectious diseases) have been proposed; see, e.g., Mollison (1995), Anderson and May (1991), Watts and May (1992), and Kretzschmar and Morris (1996). Our transmission model follows Garnett and Anderson (1994) where disease transmission is tied to the population's sexual behavior and governed by a system of differential equations. As in their model, we consider a heterogeneous population with different *risk* categories and exogenous mixing patterns (we provide precise definitions in the next section).

The studies above do not consider the effect of allocation policies on the HIV epidemic. An exception is Walensky et al. (2009), who study the effect of different ART initiation policies on life expectancy and total treatment cost in resource-limited settings but at the single-patient level; ART scarcity, however, is not explicitly considered, and the focus is on predicting patients' survival conditional on the results of a large control trial. Zaric and Brandeau (2001) study resource allocation to a number of possible *interventions* (e.g., needle exchange and condom distribution programs), when performance is measured in terms of total quality-adjusted life-years and the number of HIV infections averted: their results show that policy performance depends on factors such as incidence and prevalence. However, ART (a key HIV intervention) is not included as a possible intervention, and attention is restricted to a tractable set of allocation policies. Similarly, Long et al. (2006) develop an HIV epidemic model to study the cost-effectiveness of ART in a population of injection drug users (IDUs) and non-IDUs, but treatment allocation is not explored. Alistar et al. (2013) develop a planning tool for evaluating potential resource allocations.

Estimating PoN requires finding valid upper bounds for our MDP formulation. While ADP value function approximations might fulfill this purpose (see, e.g. Bertsekas (2007)), that is not the case of our approximation (due to constraint sampling), and unfortunately, the literature on such upper bounds is scant. Brown et al. (2010) provide a bounding technique based on information-relaxation and duality. Their technique requires solving a deterministic optimization problem along a sample path of uncertainty realization, thus it is not tractable in our setting, due to the complexity of the latter problem.

There is an abundant body of literature on ADP (see, e.g. Lai et al. (2010), Adelman and Mersereau (2008), Zhang and Adelman (2009) and references within). Our approximations are based on the LP approach to ADP (de Farias and Van Roy, 2003), with constraints sampling (de Farias and Van Roy, 2004). There has been recent interest in applying ADP to healthcare problems. For example, Maxwell et al. (2009) study ambulance redeployment via ADP, Patrick

et al. (2008) and Erdelyi and Topaloglu (2009) schedule a diagnostic facility using different approximations, and Lee et al. (2008) use the approach to examine management of dialysis therapy. To the best of our knowledge, this is the first application of ADP to optimizing treatment allocation for a stochastic and dynamic population.

While nonabandonment is unquestionably an important social and professional value, there are many ethical situations which are similar, at least in terms of their impact on individuals. For example, from the allocation prioritization rules regarding the re-transplantation of patients with scarce organs when other patients who have not been transplanted might do better (Ubel et al., 1993), and policy choices about paying for very expensive cancer drugs, policy makers are frequently confronted with the reality that "no healthcare system can provide every medical intervention that offers a prospect of health benefit to everyone, all of the time." (Faden et al., 2009). Although one may suggest that there is a difference between not starting a therapy versus discontinuing a therapy once started, some medical ethicists disagree. The General Medical Council of Britain states there is no ethical difference between withdrawing a service and not starting it (General Medical Council, 2010), and the American Medical Association (2009) notes that "There is no ethical distinction between withdrawing and withholding life-sustaining treatment." The shortage of treatment is likely to remain for the foreseeable future (Avdeeva et al., 2011; UNAIDS, 2012), thus the importance of measuring the consequences of current practice.

Medical ethicists have explored various mechanisms for allocating scarce therapies. Triage, which is defined as the "well-established process of finding the most appropriate disposition for a patient based on an assessment of the patient's illness and its urgency" (Kohlenberger et al., 1994), was recommended in the allocation of scarce therapies during high consequence events (e.g. natural disasters), where the question of whether or not withdrawing therapy can be justified even in the developed world, by Kraus et al. (2007). The latter work further notes that "It is reasonable to assume that during a high-consequence event there will be victims who are at greater risk for (salvageable) loss of life or limb than those patients already occupying inpatient beds. In such instances patients with little likelihood of medical benefit (i.e., patients for whom medical care is futile) may be the first to be considered for withheld or withdrawn care..." Prioritization for kidney transplantation in the U.S. offers a similar analogy: A few decades ago, such priorities ranked the sickest patients at the top, regardless of the prognosis, and most policy makers and healthcare providers called for research to focus on increasing the number of available organs, rather than in rationing the limited supply (Caplan, 1995); however, over time research revealed the indirect cost of such a policy, and now kidney prioritization considers post-transplant survival into account (Organ Procurement and Transplantation Network, 2014).

Using the taxonomy proposed by Persad et al. (2010), PoN can be viewed as the value of moving from prioritarianism (which prioritizes the sickest) to utilitarianism (which maximizes the societal benefits of the scarce therapies). Persad et al. (2010) do not recommend a "sickest-first" allocation, noting that "...when interventions are persistently scarce, saving the progressively ill person later will always involve depriving others. When we cannot save everyone, saving the sickest first is inherently flawed and inconsistent with the code idea of priority to the worst-off." In contrast, they do recommend a utilitarian approach, observing that "...saving more lives should be part of a multiprinciple allocation system," and "prognosis is undeniably relevant...", although they further recommend an explicit preference for the young, which our models do not consider. However, our models could be extended to fit "The Complete Lives System," the ethical framework they recommend.

3. Modeling the Price of Nonabandonment Policies

In this section, we present an MDP formulation for optimizing treatment allocation under different classes of admissible policies. Then, we formally define PoN as the gap in performance between the best policies within two classes.

3.1. Optimal Allocation of Scarce HIV Treatment

Model primitives and assumptions. We consider a Markovian discrete-time infinite-horizon model of the progression of an HIV epidemic within a population of heterogeneous individuals. At a broad level, the population is composed of susceptible individuals and infected patients.

Susceptible individuals are characterized by a risk index r, which evolves stochastically in time and signals a level of sexual activity. In particular, the index is associated with a distribution over number of partnerships established in a time period. Children are assumed to be sexually inactive, and are indexed by $r = 0$ (we consider mother-to-child transmission; see our description of transition dynamics below). We let R denote the set of risk categories, which we assume is finite.

In addition to their risk, infected patients are also characterized by treatment (p) and health (h) indices. With respect to treatment, we assume that patients start treatment at most once (we relax this assumption in Section 6), and let P denote the finite set of possible *treatment phases* an infected patient may be on at any time (we provide a precise definition of this set in Section 3.2). The health index h describes a patient's current CD4 count (a measure of the strength of the immune system that is the *primary* health indicator in practice) or a set of indicators of the evolution of her CD4 count. (Our model discretizes the range of CD4 count into the five categories commonly used in clinical practice (WHO, 2012): See Section 5 for more details.) Consequently, the index also characterizes a patient's mortality rate and quality of life.

- Susceptible individuals are in a unique representative health state. These individuals have a normal CD4 count (i.e. their CD4 count is in the highest category).
- For untreated patients, h represents their current CD4 count category.
- For patients being treated, h is a triplet indicating their: (i) CD4 count category at the beginning of treatment; (ii) maximum CD4 count category while on treatment; and (iii) current CD4 count category. This triplet allows us capture the development of resistance within a Markovian setting.

We let H denote the (finite) set of health indices. Recall that in practice the development of resistance is not observable; it is inferred using the same type of information our model considers for making treatment allocation decisions. Thus, our model does not need to explicitly consider inference of resistance (it is already embedded).

Let X_r^t and $Y_{i,p}^t$ denote the number of susceptible individuals of risk $r \in R$ and that of infected individuals in infected state $i \in I := R \times H$ and treatment phase $p \in P$, at the beginning of period t, respectively. Define $S^t := (X^t, Y^t)$ as the state of the system at the beginning of period t, where $X^t := (X_r^t : r \in R)$, and $Y^t := (Y_{i,p}^t : i \in I, p \in P)$, and let $\mathcal{S}$ denote the (unbounded) state space, i.e.,

$$\mathcal{S} := \left\{ S = (X, Y) \in \mathbb{Z}_+^{|R||I||P|} \right\}.$$

We assume that each patient undergoing treatment consumes one dose of therapy per period, and that there are K doses available per period (we assume K is constant, but relax this in Section 7). We assume that doses cannot be stored, which is in line with practice in resource-limited settings. In our model, treatment allocation decisions (*actions*) move patients from one treatment phase to another, e.g., from pre-treatment to the first treatment phase. We let $A(S)$ denote the set of actions available to policy makers in state $S \in \mathcal{S}$, and $\Gamma_a(S) \in \mathcal{S}$ the state of the system immediately after applying action $a \in A(S)$. In Section 3.2, we explain in detail the set of admissible actions for the classes of nonabandonment and abandonment-permitting policies.

Transition dynamics. In our model, an adult's risk changes within a period from r to r' with probability $\eta_{r,r'}$ for $(r, r') \in R^2$, and a child becomes an adult of risk $r \in R$ with probability η_r (i.e., transition to adulthood follows a geometric distribution). Patients in infected state i and treatment phase $p \in P$ transition to state i' within a period with probability $\rho_{i,i'}^p$ for $(i, i') \in I^2$. At each period, an adult of risk $r \in R$ gives birth to a child with probability γ_r. We model mother-to-child transmission as follows: a child born from a patient in state $(i, p) \in I \times P$ is born infected with probability $\kappa_{i,p}$. Susceptible and infected individuals of risk $r \in R$ and in state $(i, p) \in I \times P$ die within a period with probabilities d_r and $d_{i,p}$, respectively. We assume all these events are independent of each other.

Let $p_r(S,a)$ be the probability that a susceptible individual in risk category $r \in R$ becomes infected within a period when in state $S \in \mathcal{S}$ and action a is taken. Here, we assume that each susceptible individual becomes infected in different partnerships independent of each other. In Appendix B we provide error bounds for this approximation, which we show are relatively small in the settings of interest. Following Garnett and Anderson (1994),

$$p_r(S,a) := \mu_r \sum_{r' \in R} \omega_{r,r'} \frac{\sum_{h \in H} \sum_{p \in P} \nu_{r',h,p} \ \Gamma_a(S)_{i,p}}{N_{r'}(S)}, \quad S \in \mathcal{S}, \tag{1}$$

where μ_r denotes the average number of partners a susceptible individual of risk r establishes each period, $(\omega_{r,r'} : (r,r') \in R^2)$ denotes the *mixing matrix* describing partnership patterns within the population, $\nu_{r,h,p}$ represents the transmission probability in a partnership with a patient in infected state $(r,h) \in I$ undergoing treatment phase $p \in P$, $N_r(S)$ denotes the total population in $S \in \mathcal{S}$ of risk $r \in R$, and $\Gamma_a(S)_{i,p}$ denotes the number of infected patients in state (i,p) in $\Gamma_a(S)$.

In broad terms, when action $a \in A(S)$ is applied to state $S^t \in \mathcal{S}$, the state of the system immediately changes to $\Gamma_a(S^t)$, and by the beginning of period $t+1$ it is

$$S^{t+1} = \Gamma_a(S^t) + \text{Births} - \text{Deaths} + \Delta \text{ Risk transitions} + \Delta \text{ Health transitions} + \Delta \text{ HIV transmission},$$

where the terms on the right-hand side above represent vectors of correlated random variables whose distributions depend on the parameters summarized in Table 1, and that account for changes in the population within a period.

Table 1 Summary of notation

R	Set of values for risk index	$\rho^p_{i,i'}$	Transition prob. from state i to i' in treatment phase p
H	Set of values for health index	$p_r(S,a)$	Infection prob. for an individual of risk r in state S
P	Treatment phases	$\omega_{r,r'}$	Partnership prob. between individuals of risks r and r'
I	Infected state $\{(r,h)\}$	$\nu_{r,h,p}$	Infection prob. in a partnership with a patient in state (r,h,p)
K	Number of treatment doses	γ_r	Birth prob. for an individual of risk r
c	Distribution over state space	$\kappa_{i,p}$	Mother-to-child transmission prob. for patients in state (i,p)
λ	Discount factor	$\eta_{r,r'}$	Transition prob. from risk r to r'
m	Number of sampled states	μ_r	Mean partnerships for an individual of risk r
ϕ	Sampling distribution	$d_{i,p}$	Death prob. for a patient in state (i,p)
π	A stationary allocation policy	d_r	Death prob. for an individual of risk r
Σ	Set of all permutations	$\hat{c}_{\pi,T}$	Estimate of c for policy π and simulation period of T
$Y_{i,p}$	# of patients in state (i,p)	$N_r(S)$	Total number of individuals of risk r
X_r	# of susceptible ind. of risk r	$u_{i,p}$	QALYs experienced by a patient in state (i,p)

Objective and problem formulation. We aim to maximize the expected discounted cumulative quality-adjusted life-years (QALYs) of the population. This metric captures the fact that the quality of life of patients in poor health states or undergoing treatment (and its side effects) is lower than

that of healthy untreated patients: See Glasziou et al. (1990). Let $g_a(S)$ denote the immediate reward resulting from applying action $a \in A(S)$ to state $S \in \mathcal{S}$. That is,

$$g_a(S) := \sum_{r \in R} X_r + \sum_{i \in I} \sum_{p \in P} u_{i,p} \; \Gamma_a(S)_{i,p}, \quad a \in A(S)\,,\, S \in \mathcal{S},$$

where $u_{i,p} \in (0,1]$ denotes the QALYs experienced by an infected patient in state $(i,p) \in I \times P$. Let $J_\pi(S)$ denote the expected discounted cumulative QALYs of the population when $S^0 = S$ under policy $\pi \in \mathcal{P}$, where $\mathcal{P}$ denotes the set of all stationary non-anticipative policies. That is,

$$J_\pi(S) := \mathbb{E}\left\{ \sum_{t=0}^{\infty} \lambda^t \; g_{\pi_t(S^t)}(S^t) | S^0 = S \right\}, \quad S \in \mathcal{S},\, \pi \in \mathcal{P},$$

where $\pi_t(S^t)$ denotes the action selected by an admissible policy π in state S^t at time t, and $\lambda \in (0,1)$ is a discount factor. From the above, policy makers solve

$$V_\Pi(S^0) := \sup_{\pi \in \Pi} \left\{ J_\pi(S^0) \right\},$$

where $\Pi \subseteq \mathcal{P}$ denotes the set of admissible policies under consideration. To ensure that the value function V_Π is well defined, we assume that rewards are discounted at a rate larger than that of the population growth. We formalize this in Assumption 1, which we assume holds throughout the paper (we comment on the validity of this assumption in Section 5).

Assumption 1 *The discount factor $\lambda \in (0,1)$ is such that $\frac{1-\lambda}{\lambda} > \sum_{r \neq 0} \gamma_r - \min_{(i,p) \in I \times P} \{d_{i,p}\} \wedge \min_{r \in R} \{d_r\}$.*

For any real-valued function $J : \mathcal{S} \to \mathbb{R}$, define the *per-capita* supremum norm

$$\|J\|_{pc} := \sup_{S \in \mathcal{S}} \left\{ \frac{|J(S)|}{|S|+1} \right\},$$

where $|S|$ denotes the size of the population at state $S \in \mathcal{S}$. Lemma 1 below characterizes the optimal value function V_Π and an optimal policy π^*. Its proof can be found in Appendix A.

Lemma 1 *There exist $\overline{v} > 0$ and V_Π such that $\|V_\Pi\|_{pc} < \overline{v}$ and V_Π is the unique solution to*

$$J(S) = \max_{a \in A(S)} \left\{ g_a(S) + \lambda \; \mathbb{E}_a \left\{ J(S^1) | S^0 = S \right\} \right\}, \quad S \in \mathcal{S}, \tag{2}$$

such that $\|J\|_{pc} < \infty$. In addition, there exists an optimal stationary policy $\pi^* \in \Pi$, where

$$\pi^*(S) \in \arg\max_{a \in A(S)} \left\{ g_a(S) + \lambda \; \mathbb{E}_a \left\{ V_\Pi(S^1) | S^0 = S \right\} \right\}, \quad S \in \mathcal{S}.$$

In the above, $\mathbb{E}_a\{\cdot\}$ denotes expectation when action $a \in A(S)$ is selected. Note that (2) has countably many elements, thus an exact solution is unlikely to be found in general.

3.2. Price of Nonabandonment

In defining classes of admissible policies we assume that only one ART regimen is available: Secondary ART regimes are considerably more expensive (Long et al., 2010), thus we claim that scarce resources should focus on acquiring first-line regimens. To test this claim, we study the effect of allocating second-line regimens in Section 6.

For the case of abandonment-permitting policies, we restrict attention to policies that remove patients from treatment at most once, and do not reinstate treatment to patients to whom treatment was removed from in the past. We claim that such assumption is mild because in resource-poor settings there are always patients that have not been treated in the past, that benefit as much or potentially more from treatment than those how have been treated and might have developed resistance. We test this claim in Section 6, where we relax this assumption.

Observe that our model does not impose any conditions for discontinuing treatment so it is possible, for example, that ART is discontinued for a patient only after one period. In Section 6 we consider a model where discontinuing treatment is possible only for patients who have stopped responding to treatment.

With the above, the possible treatment phases in P are pre-treatment phase ($p=0$), on-treatment phase ($p=1$), and post-treatment phase ($p=2$), and

$$A(S) := \left\{ a_{i,p} \in \mathbb{Z}_+^{|I||P|} \ : \sum_{i\in I}(a_{i,1} - a_{i,2} + Y_{i,1}) \le K,\ a_{i,p} \le Y_{i,p-1}, \quad i \in I,\, p \in P \right\}, \quad S \in \mathcal{S},$$

where $a_{i,0} := 0$, $a_{i,1}$ represents the number of pre-treatment patients in infected state i moved to the on-treatment phase, and $a_{i,2}$ represents the number of on-treatment patients in infected state i moved to the post-treatment phase. With this, one has that if $S=(X,Y)$, then $\Gamma_a(S)=(X,Y'(a))$, with $Y'_{i,p}(a) = Y_{i,p} + a_{i,p} - a_{i,p+1}$, $(i,p) \in I \times P$. Note that in this setting $a \in A(S)$ corresponds to treatment initiation and termination decisions, and hence these decisions are endogenously determined by our model: See Section 7.2 for further discussion.

The class of admissible non-anticipative policies in the nonabandonment case is $\Pi^{na} := \{\pi \in \mathcal{P} : \pi(S) \in A(S);\, a_{i,2} = 0\}$. Similarly, for the abandonment-permitting case the class of admissible non-anticipative policies is $\Pi^{ap} := \{\pi \in \mathcal{P} : \pi(S) \in A(S)\}$. We define the price of nonabandonment

$$\text{PoN} := V_{\Pi^{ap}}(S^0) - V_{\Pi^{na}}(S^0).$$

PoN measures the benefit that society might obtain in terms of total QALYs by departing from the nonabandonment practice and allowing premature treatment termination. In other words, PoN represents the price that society is paying by restricting attention to nonabandonment policies, despite their effect in the health of untreated patients and the propagation of the epidemic.

4. Approximate Solutions and Performance Guarantee

Computing PoN exactly is intractable as it involves solving MDP formulations with countable state spaces. However, one can still provide a valid lower bound on PoN by computing: (i) A lower bound on the value function in the abandonment-permitting setting; and (ii) an upper bound for that in the nonabandonment setting. Next, we use the LP approach to ADP to derive allocation policies in the abandonment-permitting case, and use their performance as lower bounds. We then compute two tractable upper bounds on the value functions using a relaxation of our formulation and a natural upper bound from the LP approach to ADP.

4.1. Lower Bound

We adapt the LP approach to ADP to compute efficient treatment allocation policies for different classes of admissible policies. Similar to the case of finite state space and bounded rewards (Puterman, 2005), one can show that $J \geq V_\Pi$ for any J such that $\|J\|_{pc} < \infty$, and

$$J(S) \geq \max_{a \in A(S)} \left\{ g_a(S) + \lambda \, \mathbb{E}_a \left\{ J(S^1) | S^0 = S \right\} \right\}, \quad S \in \mathcal{S}.$$

Consider a distribution $\{c(S) : S \in \mathcal{S}\}$ such that $c(S) > 0$ for all $S \in \mathcal{S}$. Proposition 1 provides an alternative characterization of the value function.

Proposition 1 *If $U \geq \overline{v}$, then V_Π is the unique solution to*

$$\begin{aligned} \min \; & \sum_{S \in \mathcal{S}} \frac{c(S)}{|S|+1} \; J(S) \\ s.t. \; & J(S) \geq g_a(S) + \lambda \, \mathbb{E}_a \left\{ J(S^1) | S^0 = S \right\}, \quad a \in A(S), \, S \in \mathcal{S}, && \text{(3a)} \\ & |J(S)| \leq (|S|+1) \; U, \quad S \in \mathcal{S}. && \text{(3b)} \end{aligned}$$

Formulation (3) has a countable number of constraints and variables. Note that (3b) guarantees that $\|V_\Pi\|_{pc} < \overline{v}$. We approximate the solution to (3) in two steps. First, we approximate V_Π as an affine combination of basis functions to reduce the number of variables. Second, we limit the number of constraints by sampling the states that are most likely visited by the optimal policy. We also use structural properties of the "optimal" policy under such an approximation to limit the number of actions worthy of consideration (note that $A(S)$ grows exponentially with the size of S).

Value function approximation. We approximate V_Π as an affine combination of basis functions. In particular, for a vector $\alpha := \left(\alpha_0, \, (\alpha_r : r \in R), \, \left(\alpha_{(i,p)} : (i,p) \in I \times P\right)\right)$, define

$$\widetilde{V}_\alpha(S) := \alpha_0 + \sum_{r \in R} \alpha_r \; X_r + \sum_{i \in I} \sum_{p \in P} \alpha_{i,p} \; Y_{i,p}, \quad S \in \mathcal{S}, \tag{4}$$

where α_r represents a marginal change in the expected QALYs of the population if one susceptible individual of risk $r \in R$ is added to the population. A similar interpretation holds for the $\alpha_{i,p}$'s. Our first approximation restricts attention to feasible solutions to (3) of the form $\widetilde{V}_\alpha$. (Our numerical experiments also considered second order approximations that included terms of the form $X_r \cdot Y_{i,p}$. The improvement over the affine approximation was marginal, thus we omit such an analysis.) In this regard, note that $|\widetilde{V}_\alpha(S)| \le (|S|+1)U$ for all $S \in \mathcal{S}$ is equivalent to $\|\alpha\|_\infty \le U$, where $\|\cdot\|_\infty$ denotes the uniform norm. By considering the affine approximation above, the action maximizing the right-hand side of (3a) is a *state-dependent priority rule*.

Proposition 2 *For any vector α and state $S \in \mathcal{S}$, the action maximizing the right-hand side of* (3a)*, when the value function is replaced by the approximation in* (4)*, is that assigning treatment according to the priority rule induced by the ranking*

$$\sigma_{i,p}(\alpha, S) := (-1)^p \Bigg(\alpha_{i,p} d_{i,p} - \alpha_{i,p-1} d_{i,p-1} - \sum_{i' \in I} (\alpha_{i',p} \rho^p_{i,i'} - \alpha_{i',p-1} \rho^{p-1}_{i,i'}) + \frac{\sum_{r'} \alpha_{r'} X_{r'} \mu_{r'} \omega_{r',r}}{N_r(S)} (\nu_{r,h,p} - \nu_{r,h,p-1}) - (\alpha_{i,p} u_{i,p} - \alpha_{i,p-1} u_{i,p-1}) \Bigg), \quad i \in I,\, p \in P.$$

Note that σ is a permutation of elements in $I \times P$, so that for each $S \in \mathcal{S}$, the "optimal" allocation is found by considering (i,p) pairs in the order given by σ. It is worth highlighting that for a given α the priority rule induced by $\sigma(\alpha, \cdot)$: (i) Depends on the state S; and (ii) unlike WHO guidelines, it assigns priorities considering both health and risk of patients. Proposition 2 shows that, when restricting attention to the approximating architecture above, rather than considering all actions in $A(S)$, we only need to consider those induced by priority lists.

Example 1 *Suppose that $R = \{1,2\}$, $H = \{1\}$, $K = 10$, and consider abandonment-permitting policies. Recall that r, h, and p represent risk, health, and treatment phase of a patient, respectively. Table 2 shows treatment allocation for a possible state and permutation σ (e.g., patients in state $(r=1, h=1, p=1)$ have the highest priority). Following such a priority list, all patients in state $(r=1, h=1, p=1)$ receive treatment, as well as three patients in state $(r=2, h=1, p=0)$. On the other hand, treatment is terminated for four patients in state $(r=2, h=1, p=1)$, and they are moved to the post-treatment phase.*

Constraint sampling. Define Σ as the set of all permutations of elements in $I \times P$. For $\sigma \in \Sigma$ and $S \in \mathcal{S}$, define

$$Y^\sigma_{i,p}(S) := D_{i,p}(S) - d_{i,p} \left(a^\sigma_{i,p}(S) - a^\sigma_{i,p+1}(S) \right) + \sum_{i'} \rho^p_{i',i} \left(a^\sigma_{i',p}(S) - a^\sigma_{i',p+1}(S) \right), \quad i \in I,\, p \in P,$$

Table 2 Assigning treatment according to a priority list

Priority σ	r	h	p	Number of patients	Allocated treatment
1	1	1	1	4	4
2	2	1	0	3	3
3	2	1	1	7	3
4	1	1	0	2	0

$$X_r^\sigma(S) := D_r(S) - X_r\mu_r \sum_{r'}\sum_{h}\sum_{p} \frac{\omega_{r,r'}\nu_{r',h,p}}{N_{r'}(S)}(a^\sigma_{r',h,p}(S) - a^\sigma_{r',h,p+1}(S)), \quad r \in R,$$

where $a^\sigma(S) \in A(S)$ denotes the greedy ART allocation associated with permutation σ (as illustrated in Example 1), and $D_{i,p}(S)$ and $D_r(S)$ are finite state-dependent constants, independent of σ. By restricting attention to priority rule policies and imposing constraints (3a) for a finite set of m states, $\overline{\mathcal{S}}$, one can write the approximation problem as

$$RLP(c,\overline{\mathcal{S}}): \min \sum_{S\in\mathcal{S}} \frac{c(S)}{|S|+1}\left(\alpha_0 + \sum_{r\in R}\alpha_r\ X_r + \sum_{i\in I}\sum_{p\in P}\alpha_{i,p}\ Y_{i,p}\right) \tag{5a}$$

$$s.t.\ D(S) + (1-\lambda)\alpha_0 + \sum_{i\in I}\sum_{p\in P}\alpha_{i,p}\left(Y_{i,p} - \lambda Y^\sigma_{i,p}(S)\right) + \sum_{r\in R}\alpha_r\left(X_r - \lambda X^\sigma_r(S)\right) \geq 0, \quad \sigma\in\Sigma,\ S\in\overline{\mathcal{S}}, \tag{5b}$$

$$\|\alpha\|_\infty \leq U, \tag{5c}$$

where $D(S)$ is independent of σ. We propose an exact algorithm to find the optimal solution of $RLP(c,\overline{\mathcal{S}})$, which we use to construct an efficient allocation policy.

Algorithmic approach. $RLP(c,\overline{\mathcal{S}})$ is a linear program with $n := |I|\,|P| + |R| + 1$ variables and $|\Sigma|\left|\overline{\mathcal{S}}\right| + 2n$ constraints. Recall that Σ contains all permutations of $I \times P$, thus $|\Sigma| = (|I|\,|P|)!$: Even for small values of I and P, $|\Sigma|$ is prohibitively large (both in the nonabandonment and abandonment-permitting cases) so one cannot explicitly formulate $RLP(c,\overline{\mathcal{S}})$.

Let $\overline{\Sigma}(S) \subseteq \Sigma$ be a set of a priori "plausible" priority lists for state $S \in \mathcal{S}$. Suppose one solves $RLP(c,\overline{\mathcal{S}})$ for a fixed c and $\overline{\mathcal{S}}$, imposing constraints (5b) only for $\sigma \in \overline{\Sigma}(S)$, all $S \in \overline{\mathcal{S}}$, and let α^* denote the optimal solution to (5) in such a situation (i.e. its *modified* version). If for each state S the priority $\sigma(\alpha^*, S)$ is included in $\overline{\Sigma}(S)$, then one concludes that α^* is optimal when one includes all constraints in $RLP(c,\overline{\mathcal{S}})$ (those associated with permutations in $\Sigma \setminus \overline{\Sigma}(S)$); on the other hand, if $\sigma(\alpha^*, S)$ is not included in $\overline{\Sigma}(S)$, then one should add $\sigma(\alpha^*, S)$ to the latter set.

This observation allows us to solve $RLP(c,\overline{\mathcal{S}})$ iteratively: we use clinically-based guidelines (such as WHO recommendations) to construct an initial set of permutations $\overline{\Sigma}(S)$ for each $S \in \overline{\mathcal{S}}$. If $\sigma(\alpha^*, S) \notin \overline{\Sigma}(S)$, where α^* solves the modified $RLP(c,\overline{\mathcal{S}})$, then we add $\sigma(\alpha^*, S)$ to $\overline{\Sigma}(S)$ and resolve. Because $|\Sigma| < \infty$, the outcome of the procedure is the solution to $RLP(c,\overline{\mathcal{S}})$, although

convergence will depend on the initial set of policies (in our numerical experiments, we find that few permutations are added prior to convergence: See Section 5.1). See Algorithm 1.

Modified $RLP(\cdot,\cdot)$ also includes a set of *empirical lower bound constraints* $\widetilde{V}_\alpha(S) \geq \widehat{V}_\alpha(S)$ for a relatively small subset of states $\widehat{\mathcal{S}} \subset \overline{\mathcal{S}}$, where $\widehat{V}_\alpha(S)$ denotes a Monte Carlo estimate of the performance of the priority rule policy induced by α starting from $S^0 = S$. The idea behind these constraints is to recover the (upper) bounding property of the approximation, derived implicitly from Proposition 1, which might be lost due to insufficient sampling of constraints. Our results in Section 5.1 indicate that these constraints, which we believe are novel in the ADP literature, significantly improve the quality of the approximation.

Algorithm 1 Solving $RLP(c,\overline{\mathcal{S}})$

For each $S \in \overline{\mathcal{S}}$ find an initial set of policies $\overline{\Sigma}(S)$.
Let α be a solution to modified $RLP(c,\overline{\mathcal{S}})$.
while $\sigma(\alpha,S) \notin \overline{\Sigma}(S)$ for all $S \in \overline{\mathcal{S}}$ **do**
  Set $\overline{\Sigma}(S) := \overline{\Sigma}(S) \cup \{\sigma(\alpha,S)\}$.
  Re-solve modified $RLP(c,\overline{\mathcal{S}})$.
Return α as an optimal solution to $RLP(c,\overline{\mathcal{S}})$.

We now discuss how to select c and $\overline{\mathcal{S}}$. de Farias and Van Roy (2004) show that c regulates the quality of the approximation and can be used to target those regions in $\mathcal{S}$ where good approximations might be needed. We are interested in regions that are more likely to be visited by the optimal policy in the short term (note that both the policy and, therefore, the regions of interest are initially unknown). We select c and $\overline{\mathcal{S}}$ so that: (i) Solving $RLP(c,\overline{\mathcal{S}})$ is tractable; and (ii) $RLP(c,\overline{\mathcal{S}})$ generates a policy that is likely to visit states that have larger weights according to c. Regarding (i), we use $\widehat{c}_{\pi,T}$, an estimate of the distribution (over $\mathcal{S}$) induced by policy π, where T denotes the simulation budget on a Monte Carlo simulation of the system (see Appendix C for details). Regarding (ii), we use the following iterative procedure: we simulate the evolution of the population under policy π^k (π^0 is exogenously given) to compute $\widehat{c}_{\pi^k,T}$ and sample m states, which we denote by $\mathcal{S}^k$; then we formulate and solve $RLP(\widehat{c}_{\pi^k,T},\mathcal{S}^k)$ to obtain an optimal solution α^{k+1} via Algorithm 1: We iterate this step until $\|\alpha^k - \alpha^{k-1}\| \leq \epsilon$. This procedure is summarized for convenience in Algorithm 2. We discuss the convergence of the procedure in Section 5.1.

Algorithm 2 Finding approximating coefficients $\alpha(m,\epsilon)$

Initialize α^0 and $\pi^0 = \sigma(\alpha^0,\cdot)$, and set $k=1$.

Estimate $\widehat{c}_{\pi^0,T}$, sample $\mathcal{S}^0$, and solve for α^1, a solution to $RLP(\widehat{c}_{\pi^0,T},\mathcal{S}^0)$.

while $\|\alpha^k - \alpha^{k-1}\| > \epsilon$ **do**

Set $\pi^k = \sigma(\alpha^k,\cdot)$, compute $\widehat{c}_{\pi^k,T}$, sample $\mathcal{S}^k$, and solve for α^{k+1}, a solution to $RLP(\widehat{c}_{\pi^k,T},\mathcal{S}^k)$.

Set $k := k+1$.

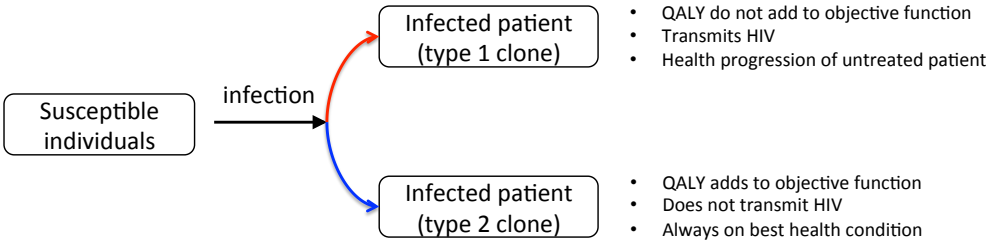
Return α^k.

4.2. Upper Bound

As noted in Section 2, finding upper bounds for MDPs is hard, and existing techniques do not apply to our setting. In this section we propose two techniques to compute upper bounds.

Clone-based upper bound. We consider a relaxation of our formulation in which each infected patient is replaced by two "clones". Type 1 clones contribute to the transmission of the disease *as if* their CD4 count category was the highest, *do not* contribute to the objective function, and exhibit health progression stochastically equal to that of patients in the pre-treatment phase. Type 2 clones *contribute* to the objective function, *do not* contribute to transmitting the disease, and their CD4 count category is the highest. Upon infection, susceptible individuals are replaced by one of each type of clone. Figure 1 illustrates the procedure. In our relaxation, we consider two sets

Figure 1 **Schematic view of the procedure for finding the clone-based upper bound**



of K treatment doses, each assigned to a different clone type. Because treatment does not affect their health progression, optimal allocation *among type 1 clones* amounts to minimizing the effects of HIV transmission. The optimal allocation is of the priority rule type.

Lemma 2 *If $d_r \le d_{i,p}$ for all r,i,p and we have random mixing, it is optimal to assign treatment to type 1 clones following the priority rule induced by $\overline{\sigma}$, where*

$$\overline{\sigma}_{r,h} := c_r(\nu_{r,h,0} - \nu_{r,h,1}) \quad r \in R,\, h \in H.$$

Likewise, as type 2 clones do not transmit HIV and their CD4 count category is the highest, treatment is *almost* irrelevant: Patients on treatment are less likely to die, so among type 2 clones it is optimal to assign all treatment doses.

The clone-based relaxation allows one to solve for the optimal treatment allocation policy by eliminating the trade-off between minimizing HIV transmission and improving the health of treated patients. Indeed, note that in the relaxed system: (i) Susceptible individuals are less likely to become infected; and (ii) infected patients are less likely to die. Thus, a sample path argument reveals that the relaxed system has more individuals than the original system in each period.

Note that by assuming that type 1 clones transmit as if their CD4 count category was the highest, the relaxation above can be seen as assuming that transmission probabilities are independent of health. In practice, however, transmission depends on the health of a patient because a higher CD4 count is indicative of lower VL. In addition, the clone-based relaxation requires random mixing. Despite the latter requirement, we use this upper bound heuristically when mixing is not random. Finally, note that the clone-based relaxation provides a unique upper bound for both the nonabandonment and abandonment-permitting settings (although treatment termination is permitted, the optimal policy is of the nonabandonment type), this implies that the bound is tighter in the abandonment-permitting setting. (We consider this fact when assessing policy performance.)

ADP-based upper bound heuristic. Recall that $\widetilde{V}_\alpha(\cdot)$ bounds the value function $V_{\Pi^{na}}(\cdot)$ from above for any α feasible for (5) when $\overline{\mathcal{S}} = \mathcal{S}$. In particular, $\widetilde{V}_{\alpha^*}(S^0) \geq V_{\Pi^{na}}(S^0)$. Unfortunately, this property does not necessarily hold when $\overline{\mathcal{S}} \subset \mathcal{S}$. However, as more states are included in $\overline{\mathcal{S}}$, the smaller the feasible region in (5) gets, which results in larger (non-decreasing) objective functions. Our selection of c aims to target regions close to S^0, thus one would expect that as more states are added to $\overline{\mathcal{S}}$, the larger $\widetilde{V}_\alpha(S^0)$ should become, eventually surpassing $V_{\Pi^{na}}(S^0)$. This observation suggests a procedure to approximate the upper bound given by $\widetilde{V}_\alpha(\cdot)$ when $\overline{\mathcal{S}} = \mathcal{S}$.

For $\epsilon > 0$ and $m \in \mathbb{Z}_+$, let $\alpha(m, \epsilon)$ denote the output of Algorithm 2 when $\left|\overline{S}\right| = m$. Based on the observation above, consider the following iterative procedure: starting from an initial sample size $m = m_0$, solve for $\alpha(m, \epsilon)$ and $\alpha(2m, \epsilon)$ and compute $\left|\widetilde{V}_{\alpha(2m,\epsilon)}(S^0) - \widetilde{V}_{\alpha(m,\epsilon)}(S^0)\right|$; if such a quantity is below the tolerance value, then we stop and consider the heuristic bound $\overline{V}(S^0) = \widetilde{V}_{\alpha(m,\epsilon)}(S^0)$; otherwise m is doubled and the procedure is repeated. Algorithm 3 formalizes this approach. The idea behind Algorithm 3 is that by increasing the size of the sample size m, the approximation from Algorithm 2 should approach a valid upper bound; note, however, that the procedure rests on the conjecture that the marginal increase in the approximation decreases as m increases. We explore the validity of such a conjecture numerically in Section 5. Note that unlike the clone-based bound, Algorithm 2 provides different bounds in the abandonment-permitting and

Algorithm 3 ADP-based upper bound

Set $m = m_0$ and compute $\alpha(m,\epsilon)$ and $\alpha(2m,\epsilon)$.
while $\left|\widetilde{V}_{\alpha(2m,\epsilon)}(S^0) - \widetilde{V}_{\alpha(m,\epsilon)}(S^0)\right| > \epsilon$ **do**
  Set $m \leftarrow 2m$ and compute $\alpha(m,\epsilon)$ and $\alpha(2m,\epsilon)$.
Return $\overline{V}(S^0) = \widetilde{V}_{\alpha(m,\epsilon)}(S^0)$.

nonabandonment settings. We will use ADP-based bounds to assess the policy performance in both settings, and compare them to clone-based bounds in the abandonment-permitting setting.

5. Numerical Study

In this section we assess policy performance, compute performance bounds, and estimate PoN using a setting mimicking current conditions in Sub-Saharan Africa. The base setting on which we test our approach considers periods with length of three months. The CD4 count range is discretized into five categories used in clinical practice: $[0,50)$, $[50,100)$, $[100,200)$, $[200,350)$ and $[350,\infty)$. This choice results in five health categories for the pre-treatment and post-treatment phases, and 55 health index triplets for on-treatment patients. (Recall that for health of on-treatment patients we consider a vector of three components, where each component is a CD4 category and has 5 possibilities; one component is maximum CD4 count and should be greater than or equal to other two components; thus we have 55 possibilities.) We estimate the disease progression probabilities between health states from the Braithwaite et al. (2011) simulation model, which is validated with data from western Kenya. Further details on the calibration of parameters can be found in Appendix F. We consider four values for the risk index of adults, differing on the average number of partnerships established in a year (thus, $|R| = 5$). Since $|P| = 3$, this results in $|\Sigma| := 300! \approx 10^{614}$ possible permutations ($|\Sigma| := 25! \approx 10^{25}$ in the nonabandonment setting). We initialize $\overline{\Sigma}(\cdot)$ by using WHO recommendations, and set $T = 75$ years, and $\left|\overline{\mathcal{S}}\right| = 1000$. Considering Assumption 1, and birth/death rates of the calibrated simulation model, we set $\lambda = 0.96$, which is aligned with values in the literature. We set the probability of becoming an adult such that the average childhood period is 15 years.

5.1. Computational Issues

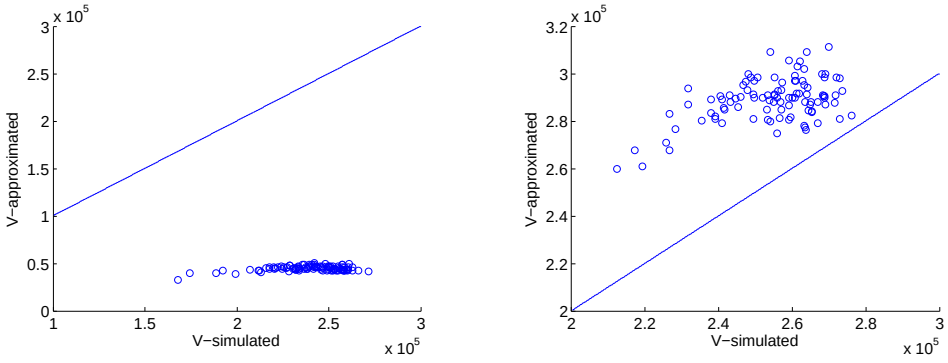
Within Algorithm 2, we consider a smooth policy update, so that $\alpha^k = \alpha^{k-1} + (\widetilde{\alpha}^k - \alpha^{k-1})k^{-3/4}$, where $\widetilde{\alpha}^k$ corresponds to the output of Algorithm 1. This type of policy update has been used in similar works to improve/guarantee convergence: See, for instance, Farias et al. (2012). (In our implementation, this smooth update is implemented after a few iterations and has little impact on the quality of the approximation.) Considering the interpretation of α as the marginal change in QALYs contributed by an individual and that $\sum_{t=0}^{\infty} \lambda^t = 25$, we set $U = 250$. We select S^0 so

that it is representative of the current state of the HIV epidemic in Sub-Saharan Africa. In our numerical experiments, we focus on maximizing life expectancy, i.e., $u_{i,p} = 1$ for all $(i,p) \in I \times P$.

In our experiments, we observe that our algorithms converge rapidly across all coverage levels: For example, in Algorithm 2, for each state $S \in \overline{\mathcal{S}}$ the algorithm adds, on average, fewer than 10 policies to $\overline{\Sigma}(S)$ prior to convergence (despite the size of $|\Sigma|$).

We illustrate the effect of empirical lower bound constraints on the quality of the value function approximation in Figure 2. The x-axis shows $\widehat{V}_{\alpha^\star}(S)$, the simulated value function under $\alpha^\star$, where $\alpha^\star$ is the optimal solution to the approximation problem. The y-axis shows $\widetilde{V}_{\alpha^\star}(S)$, which is calculated by plugging α^* into (4) for 100 states sampled randomly from $\overline{\mathcal{S}}$ (the same set of states is used in both panels). Our results show that adding the empirical lower bound constraints described in Section 4.1 significantly improves the quality of our approximations. Note that both approximate and simulated value functions differ across panels due to their dependence on α^* (which depends on whether empirical lower bounds constraints are included or not). Regarding the

Figure 2 Comparing approximated and simulated value function



(a) Without empirical lower bound constraints (b) With empirical lower bound constraints

heuristic arguments behind the ADP-based upper bound, our experiments support the hypothesis of a decreasing marginal change in the approximation. In addition, we observe that the marginal change in the approximate value function decreases to 1% after only 3 iterations of Algorithm 3, in average.

5.2. Benchmark Policies

WHO guidelines assign newly available drugs to the sickest patients (unless their CD4 count exceeds 350), and keep them on treatment until death (WHO, 2012).

State-independent priority rule policies are appealing in many practical settings, and popular among decision makers (Bertsimas et al., 2013). As a second benchmark, we consider the following two-step algorithm for finding state-independent priority rule allocation policies, for both the nonabandonment and the abandonment-permitting cases.

- First, we exploit the single-patient progression model in Braithwaite et al. (2011): we estimate the expected lifetime of a patient in infected state $i \in I$ with and without treatment via Monte Carlo, and consider the policy that prioritizes patients according to the benefit obtained by starting treatment.
- Second, we consider a two-opt algorithm that, starting from the priority list obtained by the first step above, randomly exchanges the priority of two patient types, and preserves such a change if this results in an improvement on the simulated performance of the policy. In our numerical experiments in Section 5.3, we run the procedure for 1000 iterations.

5.3. Results

We report the performance of the ADP-based policies and benchmark relative to that of WHO guidelines. Table 3 compares the performance of the ADP-based policies and benchmark for different coverage levels, and reports the ADP-based and clone-based upper bounds. Note that coverage is defined as the number of treatment doses divided by the initial infected population size. Running times are in the range of 100 to 270 minutes, in an Intel Core i5 3.30 GHz processor with 4 GB of RAM using CPLEX 11.2.

Table 3 Policy performance improvement over WHO guidelines (% of WHO guidelines)

Coverage	15	28	47	58	67	84	95
Nonabandonment state-independent benchmark	0.4	0.6	1.3	0.9	1.1	1.3	1.4
Nonabandonment ADP policy	0.4	0.6	1.3	0.9	1.3	1.4	1.5
Nonabandonment ADP-based upper bound	5.1	5.1	4.7	4.5	4.5	4.6	4.5
Abandonment-permitting state-independent benchmark	1.1	1.3	1.7	1.2	0.8	0.7	0.1
Abandonment-permitting ADP policy	14.3	14.9	14.3	12.5	11.4	10.2	8.8
Abandonment-permitting ADP-based upper bound	18.1	18.1	17.7	16.2	14.4	13.0	11.6
Clone-based upper bound	17.7	18.1	17.5	15.9	14.3	13.0	11.5

Our results show that ADP-based policies outperform WHO recommendations in the nonabandonment case, but not significantly. In addition, the small gap relative to the ADP-based upper bound suggests that the ADP-based policies are near-optimal for most coverage levels. When premature treatment termination is allowed, our results indicate that the ADP-based policy significantly improves upon WHO recommendations. Moreover, the small gap relative to the ADP-based upper bound suggests that the ADP-based policies are near-optimal for most coverage levels. In

addition, note that the clone-based upper bound essentially coincides with the ADP-based bound. This supports the validity of the latter bound.

Table 4 displays our PoN lower bound estimates for different coverage levels as a percentage of the WHO guidelines' performance. PoN lower bound estimates are at least 9% of the optimal total discounted life years of the whole population for the coverage levels observed in Sub-Saharan Africa.

Table 4 Price of nonabandonment lower bound (% of WHO guidelines)

Coverage	15	28	47	58	67	84	95
PoN lower bound	9.2	9.8	9.6	8.0	6.9	5.6	4.3

6. Sensitivity Analysis

In this section we relax some of the assumptions made in Section 3.1 and modify our model to investigate the robustness of our results.

6.1. Incorporating Second-Line Therapy

Our base model assumes that there is only one line of treatment available. Here, we incorporate a second-line therapy in our model to investigate its effect on PoN estimates.

Model extension: We accommodate the availability of a second-line therapy by adding a new treatment phase to our model in both nonabandonment and abandonment-permitting scenarios (thus, $p=2$ corresponds to the second-line therapy, and $p=3$ is the post-treatment phase). Instead of optimizing over the allocation of these doses, we assume that the policy for assigning the second-line therapy is fixed. In particular, our numerical studies assume that it follows the WHO guidelines. With this, we optimize the allocation of first-line therapy and the action space is similar to the original model. However, $\Gamma_a(S)$, the state of the system immediately after taking an action, will change due to actions related to distributing the second-line therapy. For details, see Appendix E.

Parameter calibration: The Braithwaite et al. (2011) model is able to replicate the progression of the disease in the presence of the second-line therapy. We use it to estimate the transition and death probabilities for patients in the second-line therapy phase, and assume that the patients in health status h and on the second-line therapy are as likely to transmit the disease as the patients in health status h and on the first-line therapy.

Results: We modify our optimization model and ADP algorithms to incorporate the second-line therapy and test the performance of ADP policies in both nonabandonment and abandonment-permitting settings. We re-estimate PoN to be the difference between a lower bound in the abandonment-permitting scenario and an upper bound in the nonabandonment scenario, both in

the presence of second-line therapy. (For details about performance bounds, see Appendix E.) To compare our results against scenarios where only first-line therapy is available, we associate a coverage level with the budget available to treat such a proportion of the initially infected population using only first-line therapy. Therefore, each coverage level in Table 4 is equivalent to its counterpart in Table 5 in terms of total budget. Then, assuming that second-line therapy is 2.4 times more expensive than the first-line therapy (Long et al., 2010), we assign 5%, 10%, 15%, and 20% of the total budget to the second-line therapy and report the results. Table 5 shows that PoN varies in the presence of second-line therapy, although a closer look to our results (see Table 13 in Appendix E) reveals that such a change is due mainly to poorer performance relative to when all resources are allocated to first-line therapy. Figures in Table 5 are reported as percentage of the WHO guidelines' performance when only first-line treatment is available. Our results suggest that priority should be put exclusively in the acquisition of first-line therapy.

Table 5 PoN lower bound estimates considering the second-line therapy (% of WHO guidelines)

Coverage	15	28	47	58	67	84	95
PoN base case (from Table 4)	9.2	9.8	9.6	8.0	6.9	5.6	4.3
5% of total budget for second-line therapy	9.2	9.7	9.4	7.6	5.8	5.4	4.3
10% of total budget for second-line therapy	7.9	8.5	8.8	6.9	4.7	4.5	4.0
15% of total budget for second-line therapy	8.0	8.1	7.6	6.3	4.5	4.5	3.3
20% of total budget for second-line therapy	8.0	7.9	5.6	4.5	3.8	3.2	2.8

6.2. Restricting Premature Treatment Termination

In our base model, it is possible to discontinue treatment of patients who are responding positively to treatment: our analysis in Section 7.2 shows that abandonment-permitting ADP-based policies remove treatment from patients whose CD4 counts reach 350 cells/mm^3 or start declining, which may raise additional ethical concerns. With this in mind, we examine the effect of restricting the action space so that only patients who have stopped responding to treatment are candidates for removal. For this, we modify the action set so that a patient becomes eligible for treatment termination only if her health state is deteriorating. (In our model, we impose that as far as the current health status of a patient is the same as the maximum health status during treatment, the patient stays at treatment.) We adjust our optimization framework and algorithms and test the new set of policies (see Appendix D for more details). Table 6 depicts PoN's lower bounds as a percentage of the WHO guidelines' performance. Our results indicate that the change in PoN is very limited. In addition, we observe that ADP-based policies remove patients from treatment as soon as their health state starts deteriorating.

Table 6 PoN lower bound while restricting premature treatment termination (% of WHO guidelines)

Coverage	15	28	47	58	67	84	95
PoN base case (from Table 4)	9.2	9.8	9.6	8.0	6.9	5.6	4.3
PoN lower bound	8.3	8.2	8.2	7.4	6.5	5.0	3.6

6.3. Treatment Re-initiation

In our base model, once a patients' treatment is interrupted, it cannot be continued in the future, even if the patient was responding to treatment. Here, we consider treatment re-initiation, and analyze its effects on PoN estimates. Before proceeding, we note that there is evidence that structured treatment interruptions are associated with increased risk of opportunistic infections as well as non-AIDS mortality, compared to continuous treatment (El-Sadr et al., 2006; Neaton and Grund, 2008). While in principle it is possible to incorporate such a feature in our model by expanding the set of health states H, the Braithwaite et al. (2011) simulation model does not consider such effects, thus neither do our results. Let $a_{i,3}$ denote the number of patients in infected state i that restart treatment, i.e., patients that are moved from the post-treatment phase to the on-treatment phase, and modify the action space follows:

$$A(S) := \left\{ a_{i,p} \in \mathbb{Z}_+^{|I||P|} \ : \sum_{i \in I} (a_{i,1} - a_{i,2} + a_{i,3} + Y_{i,1}) \leq K,\ a_{i,p} \leq Y_{i,p-1}, \quad i \in I, p \in P \right\}, \quad S \in \mathcal{S}.$$

Note that we do not restrict the number of times a patient may restart treatment and assume that the health progression and infectivity of a patient who restarts treatment is stochastically the same as a patient who starts treatment for the first time.

Proposition 3 *For any vector α and state $S \in \mathcal{S}$, under value function approximation* (4)*, it is optimal to (re)assign treatment according to a state-dependent priority rule.*

We modifiy our optimization model and algorithms accordingly to compute the ADP-based policies and test their performance. Table 7 displays the gain in performance when treatment re-initiation is allowed vs. when it is not, as a percentage of WHO guidelines' performace. Our results show that the policies that consider restarting treatment, slightly outperform the ones that do not. The improvement is marginal, and the changes to PoN are negligible.

Table 7 Improvement over ADP policy by considering treatment re-initiation (% of WHO guidelines)

Coverage	15	28	47	58	67	84	95
Improvement (%)	0.8	0.0	0.1	0.1	0.1	0.1	0.1

7. Case Study: HIV Simulation in Sub-Saharan Africa

In this section we develop a large-scale HIV simulation model calibrated with data from Kenya that relaxes many assumptions made in our MDP formulation. We study policy performance using the HIV simulation model and provide insights on how the ADP-based policies allocate and reallocate treatment in a variety of scenarios. For example, we evaluate how the prospect of treatment funds impacts our estimate of PoN.

7.1. HIV Simulation Model

The HIV simulation model is based on validated HIV transmission and progression models: It extends Braithwaite et al. (2011) by considering a population of patients concurrently, and incorporates the transmission dynamics in Garnett and Anderson (1994). The resulting model is able to replicate treatment scarcity and to accommodate different allocation policies.

The HIV simulation model relaxes many assumptions made in Section 3.1. In this model, HIV progression is governed by a complex biological process that depends on age, gender, CD4 count, VL, treatment history, resistance history, level of adherence, toxicity of HIV treatment, and intolerance. Thus, it not only models the evolution of the CD4 count of a patient over time, but also factors such as age and VL. Notably, it also provides QALYs estimates for patients. See Braithwaite et al. (2011) for a detailed description of the HIV progression model.

We test the performance of ADP-based policies in the setting described in Section 5. See Appendix F for further details on the calibration of this case study. Our results indicate that the ADP-based policies outperform the benchmark policies outside the MDP framework. Table 8 compares policy performance for the benchmark, relative to that of WHO guidelines. Because we do not have a valid upper bound in this more realistic setting, we cannot compute a bound on PoN. Instead, we simulate a more practical measure of PoN, which we call "simulated PoN" using the difference (relative to WHO guidelines) between the performances of abandonment-permitting and nonabandonment ADP-based policies. Note that the former policies outperform the latter for all coverage levels. In Sub-Saharan Africa, where coverage is around 50% (WHO, 2012), our estimate of the PoN bound is close to 5%. We observe that abandonment-permitting ADP-based policies sill outperform WHO guidelines in the large-scale simulation by at least 6%.

Table 8 Performance improvement over WHO guidelines' performance (% of WHO guidelines)

Coverage	15	28	47	58	67	84	95
Nonabandonment state-independent benchmark	1.2	1.3	1.5	0.5	0.1	0.9	0.3
Nonabandonment ADP-based policies	0.3	2.6	3.2	2.2	1.5	2.0	1.3
Abandonment-permitting state-indep. benchmark	0.1	1.3	2.5	1.8	1.2	1.8	1.2
Abandonment-permitting ADP-based policies	6.1	7.5	7.9	7.0	6.4	7.5	6.5
Simulated PoN	5.8	4.9	4.7	4.8	4.9	5.5	5.2

Table 9 Risk and health priorities in nonabandonment ADP-based policies

Coverage \ Prevalence	Low [0,3%)	Medium [3%,7%)	High [7%,15%)
Low [0, 35%)	High>Low>Child Healthy>Sick>Moderate	High>Low>Child Moderate>Healthy>Sick	High>Low>Child Moderate>Healthy>Sick
Medium [35, 70%)	High>Low>Child Moderate>Healthy>Sick	High>Low>Child Moderate>Sick>Healthy	High>Low>Child Healthy>Sick>Moderate
High [70%, 100%)	High>Child>Low Moderate>Sick>Healthy	High>Low>Child Healthy>Moderate>Sick	Low>Child>High Sick>Moderate>Healthy

The HIV literature shows that policy performance depends on prevalence levels. The latter, however, must be inferred from available data as the number of people infected with the disease is not known with certainty (thus neither is the actual coverage). To model disruptions arising from biased prevalence estimates, we consider a setting with a nominal coverage of 47% (that is, under the assumption of 5% prevalence; these are levels observed in Sub-Saharan Africa (WHO, 2012)), and let actual prevalence vary from 1% to 12%. Our results show that in these scenarios, PoN estimates varies between 4.4% to 8.1%.

7.2. Policy Insights

Table 9 illustrates how ADP-based policies assign priorities for treatment initiation depending on the state of the population (summarized in terms of coverage and prevalence) in the nonabandonment setting. This table summarizes the risk prioritization of ADP-based policies (recall that risk groups differ in the mean number of partnerships per period: We say a group has higher risk if it has a higher number of partnerships). For the case of health, such order is relative to current CD4 count. Note that both risk and health affect priorities: In most scenarios, riskier patients have the highest priority for receiving treatment. Also, in most scenarios, patients having moderate health are prioritized. Overall, our analyses indicate that risk is more important than health.

Table 10 provides further details on aggregated priorities under current state of the HIV epidemic: Medium coverage (50%) and medium prevalence (5%). Each entry depicts the ranking associated with the "risk-health" pair in the priority list of the ADP policy, for patients in the pre-treatment phase. (We *heuristically* sum the priorities in each column and row to find aggregate priority for each health and risk category, respectively.)

In broad terms, we see that the ADP-based policy prioritizes treatment of riskier patients, and that of those with moderate health. This order differs from those prescribed by the state-independent benchmark and WHO guidelines (the latter prioritizes the sickest patients). Note that unlike WHO policies, the ADP-based policy does not prioritize patients based solely on their health condition, as their risk is also considered. Moreover, at a more disaggregated level, the relative priorities of patients varies across risk levels.

Table 10 Aggregated priorities for health and risk in nonabandonment ADP-based policies

	Health (CD4 categories)					
	[0,50)	[50,100)	[100,200)	[200,350)	[350,∞)	
Risk (number of partners)						Aggregated risk
0	17	14	23	16	20	90
1	9	5	19	21	22	76
3	15	24	8	3	25	75
5	18	13	7	11	6	55
7	12	4	2	1	10	29
Aggregated health	71	60	59	52	83	

We also investigate the ADP-based policies' treatment-removal timing: Loosely speaking, ADP-based policies consistently stop treating a patient once her CD4 count exceeds 350 cells/mm^3, or starts declining. To check the robustness of our findings with respect to our CD4 count discretization, we separate the $[350, \infty)$ category in two subcategories: $[350, 500)$ and $[500, \infty)$, and studied treatment discontinuation times in this new setting. Our results show that treatment-removal timing is robust to this modification, i.e., abandonment-permitting ADP policies still discontinue treatment at the CD4 count of 350 cells/mm^3 or upon health deterioration.

We also compare mean treatment times: Those observed under ADP-based policies are consistently around 30% of those following WHO guidelines. As a consequence, even for low coverage levels, the ADP-based policies treat (at some point) more than 85% of the infected population.

Considering that international funds to fight the HIV epidemic have increased during the past decade, we consider a setting in which the number of treatment doses increases according to the trend of ART growth in Sub-Saharan Africa: Starting from an initial coverage of 47% we test the performance of the ADP-based policies against the WHO recommendations via the HIV large-scale simulation model. Our analysis shows that the ADP-based policies (computed for a fixed number of doses) outperform the WHO recommendations by 1%. This, and the results from the previous section, suggest that increasing coverage might mitigate the opportunity cost associated with nonabandonment practices.

8. Conclusions

After developing resistance to treatment, HIV-infected patients marginally benefit from treatment. In resource-limited settings, nonabandonment policies, which keep patients on treatment until they die, prevent the reassignment of treatment to patients who might benefit more. The current drug allocation paradigm, which arose in the developed world, avoids removing a patient from treatment. However, this view ignores the fact that in the developing world, many untreated patients might greatly benefit (relatively to patients who have developed resistance) from treatment. By removing patients from treatment, one can treat a much greater proportion of the infected population, and this results in more QALYs for the whole population.

This paper quantifies the value of moving from current prioritarianism practice (which prioritizes the sickest) to an utilitarianism approach to treatment allocation (which maximizes the societal benefits of the scarce therapies). For this, we use ADP to approximate the solution to an MDP formulation of the allocation problem. Our upper bounds indicate that the policies produced by our method are near-optimal for many settings of interest. In the context of our MDP formulation, our results show that the PoN is substantial in settings representative of the current HIV epidemic. Our sensitivity analysis suggests our results are robust with respect to the key assumptions made in the model. When tested under a large-scale simulation model that relaxes many assumptions made by the MDP formulation, our results suggest that PoN lies between 4.4% and 8.1% depending on the actual prevalence level, in settings observed in Sub-Saharan Africa. In assigning limited HIV treatment, ADP policies consider both risk and health of patients. ADP policies prioritize riskier patients and those with moderate health. This is in contrast to WHO guidelines, which recommend prioritizing the sickest patient and do not consider risk. Our results also suggest that the advantage of abandonment-permitting policies emanates mainly from their ability to reallocate therapy away from the sickest patients toward healthier and risky patients. In discontinuing treatment, ADP policies remove treatment from a patient when her CD4 count reaches the best health state or starts deteriorating.

The full ethical implications of our findings are unclear and outside the scope of this work. Nonetheless, they should contribute to addressing the difficult ethical questions that arise when departing from nonabandonment practices. In this regard, our results should inform policy makers about the opportunity cost associated with preserving nonabandonment practices, and how the magnitude of such a cost depends on, for example, the prospect of treatment funds.

Acknowledgments

The authors thank Stephen Graves, three anonymous referees, and two anonymous clinicians for their constructive feedback. We are particularly thankful to the associate editor, who suggested the analogy to the recent change in organ allocation policies.

References

Adelman, D., A.J. Mersereau. 2008. Relaxations of weakly coupled stochastic dynamic programs. *Operations Research* **56**(3) 712–727.

AIDS.gov. 2013. Global AIDS overview. Available from http://www.aids.gov/federal-resources/around-the-world/global-aids-overview/.

Alistar, S.S., M.L. Brandeau, E.J. Beck. 2013. Reach: A practical HIV resource allocation tool for decision makers. *Operations Research and Health Care Policy*. Springer, 201–223.

American Medical Association. 2009. Withholding or withdrawing life-sustaining medical treatment. Available from http://www.ama-assn.org//ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion220.page.

Anderson, R.M., R.M. May. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Science Publications, Oxford.

Avdeeva, O., J. Lazarus, M. Abdel Aziz, R. Atun. 2011. The global fund's resource allocation decisions for HIV programs: Addressing those in need. *Journal of the International AIDS Society* **14**(1).

Badri, M., S. Cleary, G. Maartens, J. Pitt, L.G. Bekker, C. Orrell, R. Wood. 2006. When to initiate highly active antiretroviral therapy in Sub-Saharan Africa? A South African cost-effectiveness study. *Antiviral Therapy* **11**(1) 63–72.

Bellan, S.E., K.J. Fiorella, D.Y. Melesse, W.M. Getz, B.G. Williams, J. Dushoff. 2013. Extra-couple HIV transmission in Sub-Saharan Africa: A mathematical modelling study of survey data. *The Lancet* **6736**(12) 61960–61966.

Bertsekas, D.P. 2007. *Dynamic Programming and Optimal Control*, vol. 2. Athena Scientific, Belmont, MA.

Bertsimas, D., V. Farias, N. Trichakis. 2013. Fairness, efficiency and flexibility in organ allocation for kidney transplantation. *Operations Research* **61**(1) 73–87.

Braithwaite, R.S., K.A. Nucifora, C.T. Yiannoutsos, B. Musick, S. Kimaiyo, L. Diero, M.C. Bacon, K. Wools-Kaloustian. 2011. Alternative antiretroviral monitoring strategies for HIV-infected patients in east Africa: Opportunities to save more lives? *Journal of International AIDS Society* **14**(30) doi: 10.1186/1758–2652–14–38.

Brown, D.B., J.E. Smith, P. Sun. 2010. Information relaxations and duality in stochastic dynamic programs. *Operations Research* **58**(4) 785–801.

Caplan, A. 1995. Organ procurement and transplantation: Ethical and practical issues. Center for Bioethics, University of Pennsylvania School of Medicine, Volume 2, Number 5; Available from http://optn.transplant.hrsa.gov/ContentDocuments/KAS_FAQs.pdf.

CIA. 2010. The World Factbook. Available from http://www.umsl.edu/services/govdocs/wofact2008.

Ciaranello, A.L., G.R. Seage, K.A. Freedberg, M.C. Weinstein, S. Lockman, R. Walensky. 2008. Antiretroviral drugs for preventing mother-to-child transmission of HIV in Sub-Saharan Africa: Balancing efficacy and infant toxicity. *AIDS* **22**(17) 2359–2369.

Cohen, M.S., C.L. Gay. 2010. Treatment to prevent transmission of HIV-1. *Clinical Infectious Diseases* **50**(Supplement 3) S85–S95.

de Farias, D.P., B. Van Roy. 2003. The linear programming approach to approximate dynamic programming. *Operations Research* **51**(6) 850–865.

de Farias, D.P., B. Van Roy. 2004. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research* **29**(3) 462–478.

El-Sadr, W.M., J.D. Lundgren, J.D. Neaton, F. Gordin, D. Abrams, R.C. Arduino, A. Babiker, W. Burman, N. Clumeck, C.J. Cohen. 2006. CD4+ count-guided interruption of antiretroviral treatment. *New England Journal on Medicine* **355**(22) 2283–2296.

Erdelyi, A., H. Topaloglu. 2009. Computing protection level policies for dynamic capacity allocation problems by using stochastic approximation methods. *IIE Transactions* **41**(6) 498 – 510.

Faden, R.R., K. Chalkidou, J. Appleby, H.R. Waters, J.P. Leider. 2009. Expensive cancer drugs: A comparison between the United States and the United Kingdom. *Milbank Quarterly* **87**(4) 789–819.

Farias, V., D. Saure, G. Weintraub. 2012. An approximate dynamic programming approach to solving dynamic oligopoly models. *RAND Journal of Economics* **43**(2) 253–282.

Fraser, C., T.D. Hollingsworth, R. Chapman, F. de Wolf, W.P. Hanage. 2007. Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proceedings of the National Academy of Sciences of the USA* **104**(44) 17441–17446.

Garnett, G.P., R.M. Anderson. 1994. Balancing sexual partnerships in an age and activity stratified model of HIV transmission in heterosexual populations. *IMA Journal of Mathematics Applied in Medicine and Biology* **11**(3) 161–192.

Garnett, G.P., R.M. Anderson. 1996. Sexually transmitted diseases and sexual behavior: Insights from mathematical models. *The Journal of Infectious Diseases* **174**(Supplement 2) 150–160.

General Medical Council. 2010. Treatment and care towards the end of life: Good practice in decision making. Available from http://www.gmc-uk.org/guidance/ethical_guidance/end_of_life_care.asp.

Glasziou, P.P., R.J. Simes, R.D. Gelber. 1990. Quality adjusted survival analysis. *Statistics in Medicine* **9**(11) 1259–1276.

Joint United Nations Programme on HIV/AIDS. 2010. UNAIDS report on the global AIDS epidemic. Available from http://www.unaids.org/en/.

Khademi, A., R.S. Braithwaite, D. Saure, A.J. Schaefer, K. Nucifora, M.S. Roberts. 2014. Should expectations about the rate of new antiretroviral drug development impact the timing of HIV treatment initiation and expectations about treatment benefits? *PLoS ONE* **9**(6) e98354.

Kirby, J. 2010. Enhancing the fairness of pandemic critical care triage. *Journal of Medical Ethics* **36**(12) 758–761.

Kohlenberger, C., C. L. Sprung, M. Danis, et al. 1994. Consensus statement on the triage of critically ill patients. *Journal of the American Medical Association* **271**(15) 1200–1203.

Kraus, C. K., F. Levy, G. D. Kelen. 2007. Lifeboat ethics: Considerations in the discharge of inpatients for the creation of hospital surge capacity. *Disaster Medicine and Public Health Preparedness* **1**(1).

Kretzschmar, M., M. Morris. 1996. Measures of concurrency in networks and the spread of infectious disease. *Mathematical Biosciences* **133**(2) 165–195.

Lai, G., F. Margot, N. Secomandi. 2010. An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. *Operations Research* **58**(3) 564–582.

Lee, C.P., G.M. Chertow, S.A. Zenios. 2008. Optimal initiation and management of dialysis therapy. *Operations Research* **56**(6) 1428–1449.

Long, E.F., M.L. Brandeau, C.M. Galvin, T. Vinichenko, S.P. Tole, A. Schwartz, G.D. Sanders, D.K. Owens. 2006. Effectiveness and cost-effectiveness of strategies to expand antiretroviral therapy in St. Petersburg, Russia. *AIDS* **20**(17) 2207–2215.

Long, L., M. Fox, I. Sanne, R. 2010. The high cost of second-line antiretroviral therapy for HIV/AIDS in South Africa. *AIDS* **24**(6) 915–919.

Maxwell, M., M. Restrepo, S.G. Henderson, H. Topaloglu. 2009. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing* **22**(2) 266–281.

Mollison, D. 1995. *Epidemic Models: Their Structure and Relation to Data*. Cambridge University Press.

Neaton, J.D., B. Grund. 2008. Earlier initiation of antiretroviral therapy in treatment-naive patients: Implications of results of treatment interruption trials. *Current Opinion in HIV and AIDS* **3**(2) 112–117.

Organ Procurement and Transplantation Network. 2014. The new kidney allocation system (KAS) frequently asked questions. Available from http://optn.transplant.hrsa.gov/ContentDocuments/KAS_FAQs.pdf.

Patrick, J., M. Puterman, M. Queyranne. 2008. Dynamic multi-priority patient scheduling for a diagnostic resource. *Operations Research* **56**(6) 1507–1525.

Persad, G.C., A. Wertheimer, E.J. Emanuel. 2010. Standing by our principles: Meaningful guidance, moral foundations, and multi-principle methodology in medical scarcity. *The American Journal of Bioethics* **10**(4) 46–48.

Phillips, A.N., A.C. Lepri, F. Lampe, M. Johnson, C.A. Sabin. 2003. When should antiretroviral therapy be started for HIV infection? Interpreting the evidence from observational studies. *AIDS* **17**(3) 1863–1869.

PLATO Collaboration. 2004. Predictors of trend in CD4-positive T-cell count and mortality among HIV-1 infected individuals with virological failure to all three antiretroviral-drug classes. *The Lancet* **364**(9428) 51–62.

Puterman, M.L. 2005. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Intersceince, New York, NY.

Revell, A.D., D. Wang, R. Wood, C. Morrow, H. Tempelman, R.L. Hamers, G. Alvarez-Uria, A. Streinu-Cercel, L. Ene. 2013. Computational models can predict response to HIV therapy without a genotype and may reduce treatment failure in different resource-limited settings. *Journal of Antimicrobial Chemotherapy* doi:10.1093/jac/dkt041.

Rottman, S.J., K.I. Shoaf, J. Schlesinger, E.K. Selski, J. Perman, K. Lamb, J. Cheng. 2010. Pandemic influenza triage in the clinical setting. *Prehospital and Disaster Medicine* **25**(02) 99–104.

Shechter, S.M., M.D. Bailey, A.J. Schaefer, M.S. Roberts. 2008. The optimal time to initiate HIV therapy under ordered health states. *Operations Research* **56**(1) 20–33.

Tabery, J., C. Mackett. 2008. The ethics of triage in the event of an influenza pandemic. *Disaster Medicine and Public Health Preparedness* **2**(2) 114–118.

Ubel, P.A., R.M. Arnold, A.L. Caplan. 1993. Rationing failure. The ethical lessons of the retransplantation of scarce vital organs. *Journal of the American Medical Association* **270**(20) 2469–2474.

UNAIDS. 2010. UNAIDS report on the global AIDS epidemic. Available from http://www.unaids.org/globalreport/Global_report.htm.

UNAIDS. 2012. AIDS dependency crisis: Sourcing African solutions. Available from http://www.unaids.org/en/resources/publications/unaidspublications/2012/.

Walensky, R.P., L.L. Wolf, R. Wood, M.O. Fofana, K.A. Freedberg, N.A. Martinson, A.D. Paltiel, X. Anglaret, M.C. Weinstein, E. Losina. 2009. When to start antiretroviral therapy in resource-limited settings. *Annals of Internal Medicine* **48**(3) 178–185.

Watts, C.H., R.M. May. 1992. The influence of concurrent partnerships on the dynamics of HIV/AIDS. *Mathematical Biosciences* **108**(1) 89–104.

Wein, L.M., S.A. Zenios, M.A. Nowak. 1997. Dynamic multidrug therapies for HIV: A control theoretic approach. *Journal of Theoretical Biology* **185**(1) 15–29.

WHO. 2010a. Antiretroviral therapy for HIV infection in adults and adolescents: Recommendations for a public health approach. Available from http://www.who.int/hiv/pub/arv/adult2010/en/index.html.

WHO. 2010b. HIV/AIDS statistics: World Health Organization. Available from http://apps.who.int/globalatlas/default.asp.

WHO. 2012. The global report for research on infectious diseases of poverty. Available from http://whqlibdoc.who.int/publications/2012/9789241564489eng.pdf.

World Bank. 2010. Life expectancy at birth, total (years). Available from http://data.worldbank.org/indicator/SP.DYN.LE00.IN/countries/ZF?display=graph,.

Zaric, G.S., M.L. Brandeau. 2001. Optimal investment in a portfolio of HIV prevention programs. *Medical Decision Making* **21**(5) 391–408.

Zhang, D., D. Adelman. 2009. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science* **43** 381–394.

Online Appendix Companion to The Price of Nonabandonment : HIV in Resource-Limited Settings

Appendix A: Proof of Results in Section 4

Proof of Lemma 1: Define $\delta := \lambda \left(\sum_{r \neq 0} \gamma_r - \min_{(i,p) \in I \times P} \{d_{i,p}\} \wedge \min_{r \in R} \{d_r\} \right)$. Assumption 1 states that $\delta \in [0,1)$. Note that $g_a(S) \leq |S| + 1$ for all $S \in \mathcal{S}$, and that

$$\lambda \, \mathbb{E}_a \left\{ (\left| S^1 \right| + 1) | S^0 = S \right\} \leq \delta(|S| + 1), \quad a \in A(S), \, S \in \mathcal{S}.$$

This implies that

$$V(S) \leq (1-\delta)^{-1}(|S| + 1), \quad S \in \mathcal{S},$$

and therefore $\|V\|_{pc} \leq \overline{v}$, where $\overline{v} := (1-\delta)^{-1}$ (Puterman, 2005, Proposition 6.10.1, p.234). Moreover, (2) has a unique solution J such that $\|J\|_{pc} < \infty$, and that π is indeed optimal (Puterman, 2005, Theorem 6.10.4, p.236). We conclude that V satisfies the Bellman equations (2).

Proof of Proposition 1: Define the dynamic programming operator T by

$$(TJ)(S) := \max_{a \in A(S)} \left\{ g_a(S) + \lambda \mathbb{E}_a \left\{ J(S^1) | S^0 = S \right\} \right\}, \quad S \in \mathcal{S}.$$

Suppose $J' \geq J$ component-wise, then $(TJ') \geq (TJ)$. In addition, Assumption 1 implies that (see proof of Lemma 1) T is a contraction in the Banach space induced by $\|\cdot\|_{pc}$ (Puterman, 2005, Theorem 6.10.4, p.236). Moreover, for any J such that $\|J\|_{pc} < \infty$ one has that

$$\lim_{n \to \infty} \|T^n J - V\|_{pc} = 0.$$

Note that V is a feasible solution to (3), and that all feasible solutions attain finite objectives. Consider a feasible J; because $J \geq TJ$ one has that $TJ \geq T^2 J \geq T^3 J \geq \ldots$. However, the sequence $T^k J$ converges to V, thus we conclude that $J \geq V$, which implies that

$$\sum_{S \in \mathcal{S}} \frac{c(S)}{|S|+1} J(S) \geq \sum_{S \in \mathcal{S}} \frac{c(S)}{|S|+1} V(S),$$

implying that V is the unique optimal solution to (3).

Proof of Proposition 2: Consider the set of constraints (3a) associated with $S \in \mathcal{S}$. We have

$$\mathbb{E}_a \left\{ X_r^1 | S^0 = S \right\} = D_r(S) - X_r \mu_r \sum_{r'} \sum_{h} \sum_{p} \frac{\omega_{r,r'} \nu_{r',h,p}}{N_{r'}} (a_{r',h,p} - a_{r',h,p+1}),$$

$$\mathbb{E}_a \left\{ Y_{i,p}^1 | S^0 = S \right\} = D_{i,p}(S) - d_{i,p}(a_{i,p} - a_{i,p+1}) + \sum_{i' \in I} \rho_{i',i}^{p} (a_{i',p} - a_{i',p+1}), \quad i \in I, \, p \in P.$$

Using $\mathbb{E}_a \left\{ Y_{i,p}^1 | S^0 = S \right\}$, and $\mathbb{E}_a \left\{ X_r^1 | S^0 = S \right\}$ one can rewrite (3a) as a single nonlinear constraint

$$\sum_{i \in I} \sum_{p \in P} \left((\alpha_{i,p} - u_{i,p}) \, Y_{i,p} - \lambda \alpha_{i,p} D_{i,p}(S) \right) + \sum_{r \in R} \left((\alpha_r - 1) \, X_r - \lambda \alpha_r D_r(S) \right) +$$

$$(1-\lambda)\alpha_0 \geq \lambda \max_{a \in A(S)} \left\{ \sum_{i \in I} \sum_{p \in P} \alpha_{i,p} \left(u_{i,p}(a_{i,p} - a_{i,p+1}) + \sum_{i' \in I} \rho_{i',i}^{p} \left(a_{i',p} - a_{i',p+1} \right) - d_{i,p} \left(a_{i,p} - a_{i,p+1} \right) \right) \right\}.$$

For a fixed α, the right-hand side above can be rewritten as

$$\max \sum_{i\in I}\sum_{p\in P}\sigma_{i,p}(\alpha,S)\ z_{i,p} \tag{A-1a}$$

$$\text{s.t.}\ \ \sum_{i\in I}(z_{i,1}+z_{i,2})\le K \tag{A-1b}$$

$$0\le z_{i,p}\le Y_{i,p-1}\quad i\in I\,,\,p=1,2. \tag{A-1c}$$

The maximization in (A-1) is a bounded knapsack problem with equal weights, therefore it is greedily solvable. Thus, for each state S, $\sigma_{i,p}(\alpha,S)$ induces a priority rule policy. This proves the abandonment-permitting case. One can use a similar argument for the nonabandonment case.

Proof of Lemma 2: In random mixing the optimal one-step policy follows from solving

$$\begin{aligned}\max\ &\sum_{r}\sum_{h}c_r(\nu_{r,h,0}-\nu_{r,h,1})a_{r,h}\\ \text{s.t.}\ \ &\sum_{r}\sum_{h}a_{r,h}\le K\\ &a_{r,h}\ge 0,\end{aligned}$$

which is a knapsack with equal weights and thus can be solved greedily. A sample path argument shows that one always prefer to have an individual as susceptible instead of infected (because $d_r\le d_{i,p}$ and infected individuals transmit the disease). Therefore, the optimal one-step policy is optimal.

Proof of Proposition 3: We modify $\mathbb{E}_a\left\{X_r^1|S^0=S\right\}$ and $\mathbb{E}_a\left\{Y_{i,p}^1|S^0=S\right\}$ as follows

$$\mathbb{E}_a\left\{X_r^1|S^0=S\right\}=D_r(S)-X_r\mu_r\sum_{r'}\sum_{h}\sum_{p}\frac{\omega_{r,r'}\nu_{r',h,p}}{N_{r'}}(a_{r',h,p}-a_{r',h,p+1})-X_r\mu_r\sum_{r'}\sum_{h}\omega_{r,r'}\frac{\nu_{r',h,1}a_{r',h,3}}{N_{r'}},$$

$$\mathbb{E}_a\left\{Y_{i,p}^1|S^0=S\right\}=D_{i,p}(S)-d_{i,p}(a_{i,p}-a_{i,p+1})+\sum_{i'\in I}\rho_{i',i}^{p}(a_{i',p}-a_{i',p+1}),\quad i\in I,\,p=0,2,$$

$$\begin{aligned}\mathbb{E}_a\left\{Y_{i,p}^1|S^0=S\right\}=&D_{i,p}(S)-d_{i,p}(a_{i,p}-a_{i,p+1})-d_{i,p}a_{i,p+2}+\sum_{i'\in I}\rho_{i',i}^{p}(a_{i',p}-a_{i',p+1})\\&+\sum_{i'\in I}\rho_{i',i}^{p}a_{i',p+2},\quad i\in I,\,p=1.\end{aligned}$$

Therefore, one can rewrite the constraint in the LP formulation as

$$\begin{aligned}&\sum_{i\in I}\sum_{p\in P}\left((\alpha_{i,p}-u_{i,p})\,Y_{i,p}-\lambda\alpha_{i,p}D_{i,p}(S)\right)+\sum_{r\in R}\left((\alpha_r-1)\,X_r-\lambda\alpha_r D_r(S)\right)+(1-\lambda)\alpha_0\ge\\&\lambda\max_{a\in A(S)}\Bigg(\sum_{i\in I}\sum_{p\in P}u_{i,p}(a_{i,p}-a_{i,p+1})+\sum_{i\in I}\sum_{p\in P}\alpha_{i,p}\Bigg(\sum_{i'\in I}\rho_{i',i}^{p}\,(a_{i',p}-a_{i',p+1})-d_{i,p}\,(a_{i,p}-a_{i,p+1})+\\&\mathbf{1}_{\{p=1\}}(\sum_{i'\in I}\rho_{i',i}^{1}a_{i',3}-d_{i,1}a_{i,3})\Bigg)+\sum_{r}\alpha_r X_r\mu_r\Bigg(\sum_{r'}\sum_{h}\sum_{p}\frac{\omega_{r,r'}\nu_{r',h,p}}{N_{r'}}(a_{r',h,p}-a_{r',h,p+1})\\&-\sum_{r'}\sum_{h}\omega_{r,r'}\frac{\nu_{h,1}a_{r',h,3}}{N_{r'}}\Bigg)\Bigg),\end{aligned}$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function. For a fixed α, the right-hand side can be written as

$$\begin{aligned}\max\ &\sum_{i\in I}\sum_{p\in P}\sigma_{i,p}(\alpha,S)\ z_{i,p}\\ \text{s.t.}\ \ &\sum_{i\in I}(z_{i,1}+z_{i,2}+z_{i,3})\le K\\ &0\le z_{i,p}\le Y_{i,p-1}\quad i\in I\,,\,p=1,2,3,\end{aligned}$$

which is a knapsack problem which induces that the optimal policy is of the state-dependent priority rule.

Appendix B: HIV Transmission Dynamics Formulation

We consider a transmission model based on Garnett and Anderson (1994), who use a deterministic compartmental framework to describe the transmission dynamics. In their model transmission is governed by a system of ordinary differential equations: The total number of susceptible individuals of risk r who become infected during a period is given by

$$\nu X_r \sum_{r'} \omega_{r,r'} \frac{Y_{r'}}{N_{r'}},$$

where ν denotes the transmission probability per partnership. For $S \in \mathcal{S}$, let $p_r(S)$ denote the probability of contagion of a susceptible individual of risk r, $A_{r,n}$ denote the event that a susceptible individual of risk r becomes infected in his/her n-th partnership, and B_r denote the (random) number of partnerships established in a period (distributed according to $\{q_r(n) : n \geq 0\}$). One has that

$$p_r(S) = 1 - \sum_{n\geq 0} \mathbb{P}\left(\bigcap_{m=1}^{n} A_{r,m}^c\right) q_r(n) = 1 - \sum_{n\geq 0} (1-\psi_r(S))^n q_r(n),$$

where we have assumed that the events $A_{r,m}$ are independent and equally likely, and $\psi_r(S)$ denotes the probability that an individual of risk r becomes infected in a partnership. Following Garnett and Anderson (1994) one has that

$$\psi_r(S) = \sum_{r'\in R} \omega_{rr'} \frac{\sum_{h,p} \nu_{r',h,p} \ (\Gamma_a(S)_{i,p})}{N_{r'}(S)},$$

where $\omega_{rr'}$ is the probability that a susceptible individual of risk r forms a partnership with someone of risk r', and the term $\sum_{h,p} \Gamma_a(S)_{i,p}/N_{r'}(S)$ represents the probability that a partnership is established with an infected patient, conditional on such a partner having risk r'. Note that $\nu_{r',h,p}$, the probability of transmission in a partnership, depends on the risk index r', health state h, and treatment phase p of the infected partner.

By using a Taylor expansion of $(1-\psi_r)^n$ (we drop the dependence on S for simplicity) we have that

$$(1-\psi_r)^n \approx 1 - n\psi_r + \mathcal{O}(\psi_r^2 \frac{n(n-1)}{2}), \tag{B-2}$$

where $\mathcal{O}$ is the error order for this approximation. By using (B-2) we approximate $p_r(S)$ by $\hat{p}_r(S)$, where

$$\hat{p}_r(S) = 1 - \sum_{n\geq 0} (1 - n\psi_r) q_r(n) = \mu_r \psi_r(S). \tag{B-3}$$

Thus, one has that

$$|p_r(S) - \hat{p}_r(S)| \leq \mathcal{O}(\frac{\psi_r^2}{2})(Var(B_r) + \mu_r^2 - \mu_r), \quad S \in \mathcal{S}.$$

In settings of interest $\psi_r < 0.01$ for states close to S^0, thus the approximation error is relatively small for plausible distributions for the number of partnerships. The exact value for ψ_r depends on the partnership patterns among individual with different risk, which is governed by a mixing matrix $\Omega = (\omega_{rr'})$: $\omega_{rr'}$ denotes the probability that a person of risk r forms a partnership with another person of risk r'. The structure of this matrix determines the mixing patterns. Two extreme cases are: (i) The assortative case, where individuals only partner with individuals of the same risk; and (ii) the disassortative case, where individuals do not partner with individuals with their same risk. An alternative setup is the so-called random mixing case in which the probabilities in the matrix are proportional to the total "supply" of sexual partnerships with

individuals of risk r. In other words, in random mixing we have that $\omega_{rr'} = \frac{N_{r'}c_{r'}}{\sum_{r''} N_{r''}c_{r''}}$, where N_r is the total number of individuals of risk r and c_r is the mean number of partnerships such individuals form in a period. Let $\Delta = [\delta_{rr'}]$ denote the $|R|$-dimensional identity matrix, and θ denote the *degree of assortative mixing*. We use the following mixing pattern, which is common in mathematical studies of STDs,

$$\omega_{rr'} = (1-\theta)\delta_{rr'} + \theta\big(\frac{N_{r'}c_{r'}}{\sum_{r''} N_{r''}c_{r''}}\big).$$

This approach is suitable for numeric simulations and captures many different mixing scenarios (Garnett and Anderson, 1994). We use $\theta = 0.7$ in the numerical experiments of Sections 5 and 7.

Appendix C: Sampling Constraints

Regarding the selection of c, one can show that minimizing (5) is equivalent to minimizing

$$\sum_{S\in\mathcal{S}} \frac{c(S)}{|S|+1}\left|\widetilde{V}_\alpha(S) - V(S)\right|.$$

As noted by de Farias and Van Roy (2003), c regulates the quality of the approximation across $\mathcal{S}$, and can therefore be used to target certain regions of the state space where one aims to obtain better approximations. In that regard, we would like to obtain better approximations in the states that are most likely to be visited in the near future when the *optimal policy* is used. For a policy π, define the distribution c_π by

$$c_\pi(S) := (1-\lambda)\sum_{t=0}^{\infty}\lambda^t \mathbb{P}_\pi\left\{S^t = S | S^0\right\}, \quad S\in\mathcal{S}.$$

We would like to use c_{π^*} in the objective function of (5), and also in sampling $\overline{\mathcal{S}}$ through ψ, as prescribed in de Farias and Van Roy (2004, Theorem 3.1., p.469). Unfortunately, one does not have prior access to π^*. Let π^0 denote an initial allocation policy. Finding c_{π^0} is computationally intractable, thus, we settle for approximating it using its empirical counterpart, $\hat{c}_{\pi^0,T}$, where

$$\hat{c}_{\pi,T}(S) := (1-\lambda)\sum_{t=0}^{T}\lambda^t \mathbf{1}_{\left\{S^t(\omega)=S\right\}}, \quad S\in\mathcal{S},\, \pi\in\mathcal{P},$$

where $\{S^t(\omega) : t\geq 0\}$ represents the outcome of a Monte Carlo simulation, and $T > m$ denotes the simulation budget. (In our numerical study we take the average over many replications.) We use this simulation run to select $\overline{\mathcal{S}}$ as well: Computing ψ is infeasible, thus we approximate it as $\hat{c}_{\pi,T}$. Note that the underlying motivation is that on most simulation runs, states are visited at most once, thus we approximate $\mathbb{P}_\pi\{S^t = S | S^0 = S'\} \approx \sum_{n=1}^{N}\mathbf{1}_{\left\{S^t(\omega_n)=S\right\}}/N$, where N is the number of replications. This results in $\psi \approx c$.

Appendix D: Restricting Premature Treatment Termination

Consider a setting where a patient becomes eligible for treatment termination only if her health state is deteriorating. The action space is represented by

$$A(S) := \Bigg\{ a_{(r,h_0,h_m,h_t,p)} \in \mathbb{Z}_+^{|I||P|} \;:\; \sum_{(r,h_0,h_m,h_t)\in I} \left(a_{(r,h_0,h_m,h_t,1)} - a_{(r,h_0,h_m,h_t,2)} + Y_{(r,h_0,h_m,h_t,1)}\right) \leq K,$$

$$a_{(r,h_0,h_m,h_t,p)} \leq Y_{(r,h_0,h_m,h_t,p-1)},\; a_{(r,h_0,h_m,h_t,1)} = 0,\; \text{for } h_t = h_m, \quad (r,h_0,h_m,h_t)\in I,\, p\in P \Bigg\}, \quad S\in\mathcal{S},$$

where h_0 denotes the initial health status, h_m denotes the maximum health status on treatment, and h_t denotes the current health status of a patient. In particular, setting $a_{(r,h_0,h_m,h_t,1)} = 0$, for $h_t = h_m$, ensures that treatment can not be terminated as long as a patient's health is at his historical best. We adjust our optimization framework and algorithms and test the new set of policies. Table 11 shows the lower bounds and upper bounds for nonabandonment and abandonment-permitting ADP policies used in Table 6 in Section 6.2, where we restrict premature treatment termination.

Table 11 Policy performance improvement over WHO guidelines in the restricted premature treatment settings (% of WHO guidelines)

Coverage	15	28	47	58	67	84	95
Nonabandonment ADP policy	0.4	0.6	1.3	0.9	1.3	1.4	1.5
Nonabandonment ADP-based upper bound	5.1	5.1	4.7	4.5	4.5	4.6	4.5
Abandonment-permitting ADP policy	13.4	13.3	12.9	11.9	11.0	9.6	8.1
Abandonment-permitting ADP-based upper bound	17.8	17.9	17.1	15.5	13.9	13.0	12.8

Appendix E: Second-Line Therapy Analysis

We incorporate second-line therapy into the nonabandonment and abandonment-permitting scenarios to re-estimate PoN.

Abandonment-permitting scenario: For the abandonment-permitting scenario, we introduce a new treatment phase, $p = 3$, which corresponds to the second-line therapy. Therefore, $Y_{i,3}$ indicates the number of patients in infected state i and under second-line therapy. Let K' denote the total number of second-line treatment doses available. We assume that the policy for assigning the second-line therapy is fixed. Therefore, the action space is similar to the original model. However, $\Gamma_a(S)$ which is the state of the system immediately after taking an action, will change due to actions related to distributing the second-line therapy. We incorporate the WHO policy regarding the second-line therapy distribution into $\Gamma_a(S) = (X, Y'(a))$. That is, initially, $Y'_{i,p} = Y_{i,p} + a_{i,p} - a_{i,p+1}$ for $p = 0, 1, 2$, and then we use the WHO policy for assigning the second-line therapy which affects $Y'_{i,2}$ and $Y'_{i,3}$. Let a'_i be the number of patients in infected state i that are moved to second-line therapy from the post-treatment phase. Note that calculating a'_i is straightforward because we fix a policy that assigns treatment to the sickest patient first subject to $\sum_i a'_i = K' - \sum_i Y_{i,3}$. In other words, we remove patients from the first-line therapy according to the actions produced by the optimization model and then assign available second-line therapies to patients who have just been removed from the first-line therapy. The evolution of the system can be represented by $Y'_{i,2}(a) = Y'_{i,2} - a'_i$, and $Y'_{i,3}(a) = Y'_{i,3} + a'_i$.

Assuming that second-line therapy is 2.4 times more expensive that first-line therapy and that policy makers are willing to spend at most half of available budget to second-line therapy, the upper bound on coverage of second-line therapy is roughly 20%. Table 12 shows the performance of ADP policies compared to that of WHO guidelines for the abandonment-permitting scenario where 10% of budget is assigned to second-line therapy. The performance of ADP-based policies is significantly better than that of WHO guidelines.

Table 12 Policy performance improvement over WHO guidelines of abandonment-permitting policies in the presence of second-line therapy (10%) as % of WHO guidelines

Coverage	15	28	47	58	67	84	95
Abandonment-permitting ADP policy	13.1	13.9	14.1	12.2	10.1	9.9	8.9
Abandonment-permitting ADP-based upper bound	18.3	18.8	18.0	16.6	16.0	14.3	13.6

Nonabandonment scenario: In the nonabandonment scenario patients receive treatment until they die. However, since there are two lines of therapy, when a patient develops resistance and the first-line therapy fails, she will start using the second-line therapy upon availability. We use WHO guidelines to determine when the first-line therapy is failed. To express the second-line therapy for the nonabandonment scenario in our model, we add a new treatment phase $p = 2$ (note that in the abandonment-permitting scenario $p = 2$ indicates the post-treatment phase and $p = 3$ indicates the second-line therapy treatment phase). Therefore, $Y_{i,2}$ denotes the number of patients in infected state i using the second-line therapy. Similar to the abandonment-permitting scenario, we assume that the policy to assign the second-line therapy follows the WHO recommendations. Here we modify $\Gamma_a(S) = (X, Y'(a))$. We know that $Y'_{i,1} = Y_{i,1} + a_{i,1}$ and $Y'_{i,0} = Y_{i,0} - a_{i,1}$. Let a'_i be the number of patients in infected state i moved from the first-line therapy to the second-line therapy. We calculate a'_i based on WHO rule subject to $\sum_i a'_i = K' - \sum_i Y_{i,2}$. The state of the system immediately after taking action is represented by $Y'_{i,1}(a) = Y'_{i,1} - a'_i$, and $Y'_{i,2}(a) = Y'_{i,2} + a'_i$. Table 13 shows the performance of ADP-based policies compared to that of WHO guidelines when 10% of budget is assigned to second-line therapy. We summarize the results of ADP policies in nonabandonment and abandonment-permitting settings in the presence of second-line therapy in Tables 14, 15, and 16 when 5%, 15% and 20% of available funds are allocated to second-line therapy, respectively.

Table 13 Policy performance improvement over WHO guidelines of nonabandonment policies in the presence of second-line therapy (10%) as % of WHO guidelines

Coverage	15	28	47	58	67	84	95
Nonabandonment ADP policy	0.2	0.5	1.2	0.9	1.2	1.2	1.0
Nonabandonment ADP-based upper bound	5.2	5.4	5.3	5.3	5.4	5.4	4.9

Table 14 Policy performance improvement over WHO guidelines (as % of WHO guidelines) in the presence of second line therapy (5%)

Coverage	15	28	47	58	67	84	95
Nonabandonment ADP policy	0.15	0.6	1.3	0.5	1.1	1.3	1.2
Nonabandonment ADP-based upper bound	4.9	4.9	4.8	4.6	4.6	4.6	4.5
Abandonment-permitting ADP policy	14.1	14.6	14.2	12.2	10.4	10.0	8.8
Abandonment-permitting ADP-based upper bound	18.3	17.7	17.5	15.7	15.7	14.0	14.3

Table 15 Policy performance improvement over WHO guidelines (as % of WHO guidelines) in the presence of second line therapy (15%)

Coverage	15	28	47	58	67	84	95
Nonabandonment ADP policy	0.3	0.6	0.9	0.7	0.6	0.9	0.2
Nonabandonment ADP-based upper bound	4.7	4.7	4.7	4.6	4.6	4.5	4.4
Abandonment-permitting ADP policy	12.7	12.8	12.3	10.9	9.1	9.0	7.7
Abandonment-permitting ADP-based upper bound	15.4	16.8	16.0	14.7	14.7	13.5	13.2

Table 16 Policy performance improvement over WHO guidelines (as % of WHO guidelines) in the presence of second line therapy (20%)

Coverage	15	28	47	58	67	84	95
Nonabandonment ADP policy	0.2	0.2	0.5	0.6	0.5	0.8	0.9
Nonabandonment ADP-based upper bound	5.0	4.4	4.7	4.5	4.3	4.5	4.6
Abandonment-permitting ADP policy	13.0	12.3	10.3	9.0	8.1	7.7	7.4
Abandonment-permitting ADP-based upper bound	16.9	17.5	16.9	14.6	12.8	13.1	12.5

Appendix F: Data and Parameter Calibration

For parameter calibration we use data from HIV literature in Sub-Saharan Africa. In particular, we use the estimates in Fraser et al. (2007) to compute the probability of transmission per sexual encounter. In addition, following Cohen and Gay (2010), we assume that patients on ART are 70% less likely to transmit the disease. The probabilities in Ciaranello et al. (2008) are used for mother-to-child transmission.

The transition probability from childhood to adulthood is such that the expected duration of childhood is 15 years (individuals older than 15 years are considered sexually active (UNAIDS, 2010)). We set the initial number of children to be 40% of the population (as almost 40% of the population are children in Sub-Saharan Africa (CIA, 2010)). Following Garnett and Anderson (1996), we use four values for the risk index for adults. Based on data from Kenya (Bellan et al., 2013), we assume adults with the first risk index constitute 89% of the population and establish (in average) one partnership per period. The remaining 11% of the population establish multiple partnerships: Adults with the second, third and fourth risk indices establish (in average) 3,5, and 7 partnerships and their population size ratio is 3:2:1, respectively. Since there is no available data on the change of individual's sexual behavior in Sub-Saharan Africa, we assume that adults do not change their risk category. The death rate for susceptible individuals is set to match life expectancy in Sub-Saharan Africa (World Bank, 2010).

For calibrating HIV transmission in the model, we compare the simulation results for the prevalence of HIV with the actual prevalence in Kenya from 1990 to 2000. HIV interventions such as condom use, education, and ART were insufficiently available in Kenya during that period, and the actual progression of the disease is observed intervention-free (UNAIDS, 2010). (After the year 2000, several interventions have been implemented concurrently in Kenya.) Figure 3 shows the results of simulation in the absence of treatment and the actual HIV prevalence in Kenya when no intervention was available.

Figure 3 Model calibration