CHAPTER 4

The Effects of Thailand's Current AIDS Policy

The evolution of Thailand's AIDS policy and the main features of its current policy for treatment (National Access to Antiretrovirals Program for People Living with AIDS, or NAPHA) are described in chapters 2 and 3. This chapter first presents a framework for analyzing the effect of government policy and then presents projections of the costs and effects of the NAPHA policy in comparison with a baseline projection of what would have happened in the absence of NAPHA. The considerable fiscal burden of the NAPHA policy is projected and compared with the total health budget and the total AIDS budget.

Framework for AIDS Policy Analysis

Governments implement policies by wielding one or more policy instruments under their control. These instruments can be grouped as follows:

- technical or engineering instruments, such as draining wetlands to combat malaria
- legal instruments, which give citizens both rights and obligations and must be enforced by a judicial system
- voluntary or behavioral instruments, which give people new information or change the prices or other incentives that guide their behavior.

Because the risky sexual and drug-using behaviors that transmit sexually transmitted diseases are private or even clandestine activities, technical and legal instruments have little force to prevent transmission. Government prevention programs have largely relied on behavioral instruments, such as providing information about the danger of HIV infection or subsidizing the prices of condoms. The discovery of a vaccine against AIDS would provide governments with a technical instrument, but inducing people to be vaccinated may require behavioral instruments.

At first glance, treatment with antiretroviral therapy (ART) may seem to be a purely technical instrument to address the damaging consequences of AIDS. However, Thailand's goals of expanding treatment to all HIV-infected residents with CD4 counts less than 200 cells per cubic millimeter and maintaining those people in good health as long as possible are ambitious. Achieving them will require the country to use all available instruments. In launching NAPHA, the government has already used the technical instrument of producing GPO-vir and the legal instrument of authorizing its production. It has mandated that government health facilities offer voluntary counseling and testing (VCT) for HIV infection, CD4 tests for those who are infected, and treatment for the eligible. By offering the treatment at nominal charges, it is also using a price instrument. The exercise of these instruments has resulted in a remarkable scale-up of treatment—from fewer than 3,000 persons living with HIV/AIDS (PHAs) under publicly financed ART in 2001 to more than 50,000 in early 2005. The challenge is to maintain those patients in good health for as long as possible.

From Policies to Epidemiology to Performance

To estimate the outcome of NAPHA and then the cost-effectiveness of various modifications of it, we must model how the specific changes introduced by this policy influence the behavior of patients and providers. Our approach is to construct a model of the links between government instruments and policy outcomes.

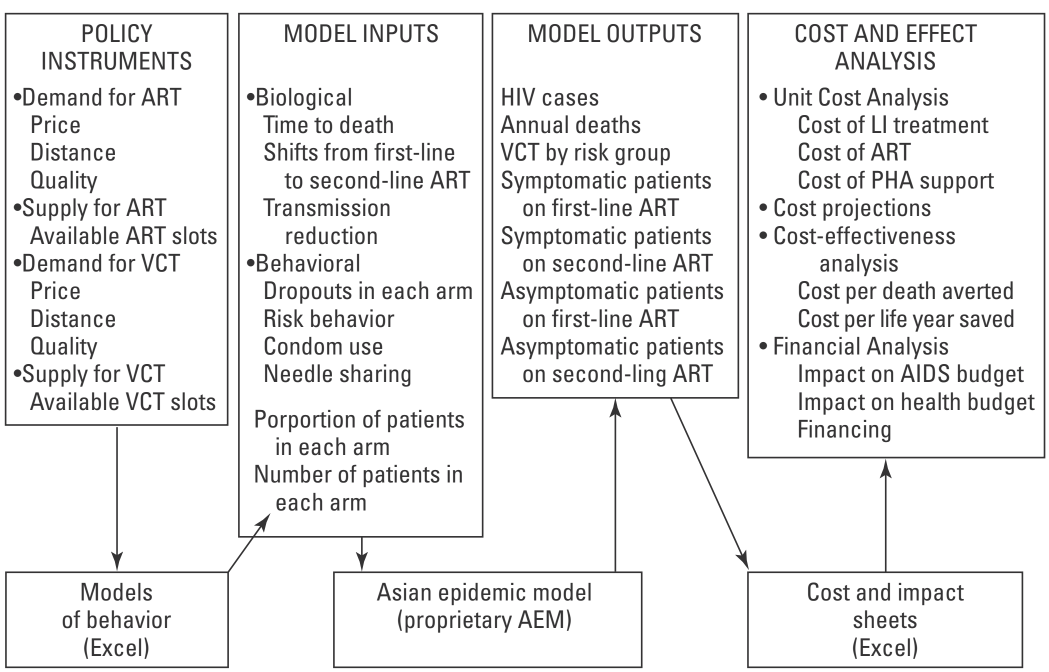
We do this in five steps (figure 4.1):

- *Step 1.* We model the link from the two primary policy instruments, price and availability (or supply), to the distribution of patient demand for care.

- *Step 2.* We project the evolution of prices and availability into the future and compute the projected distribution of demand across treatment options.
- *Step 3.* We apply the expected demand to the population of eligible infected people, as projected by an updated version of the Asian Epidemic Model (AEM), to deduce the future use of ART.
- *Step 4.* We estimate the direct and indirect health benefits of ART use.
- *Step 5.* We apply the unit costs from chapter 3 to estimate the financial burden of the NAPHA policy.

In this chapter, these five steps are applied to the NAPHA policy and to the baseline as a comparator. In chapter 5, using the same model, we apply steps 2 through 5 to three modified versions of the NAPHA policy. We then compare the stream of future health benefits with the stream of future costs for NAPHA and for three alternatives and estimate the cost-effectiveness of each version of the NAPHA policy.

Figure 4.1 From Polices to Epidemiology to Performance



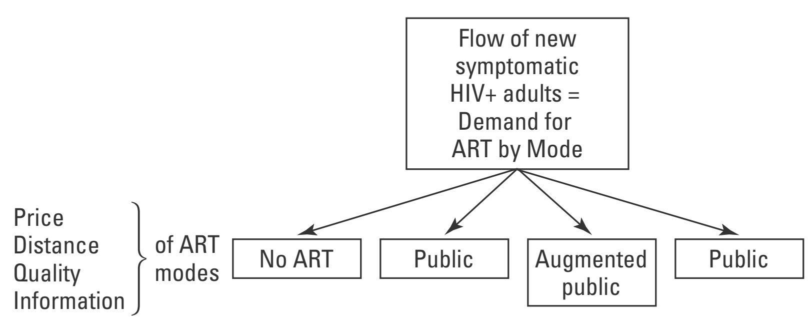
Source: Authors.
Note: OI = opportunistic infection.

Modeling the Effects of Policy Instruments

The demand for a health service is typically modeled by relating the quantity demanded to the price the patient must pay for the service, the distance he or she must travel, the quality of the care, and other relevant attributes (Akin, Guilkey, and Denton 1995; Akin and others 1998; Christianson 1976; Cissé 2004; Gertler and van der Gaag 1990; Levy and Germain 1994; Litvack and Bodart 1993; Mwabu, Ainsworth, and Nyamete 1993; Mwabu and Mwangi 1986).[1] Because a sick person's knowledge of the quality and effectiveness of a health service can also influence demand, the effect of information about quality should be modeled as well. Figure 4.2 illustrates the assumption that these four variables and others affect a symptomatic AIDS patient's decision about whether to seek treatment and which of three treatment options to choose.

It is unfortunately quite possible for some patients to choose not to seek treatment for an illness, even if they have good information. The option of no treatment is especially likely among poor people in a poor country, for whom only a few kilometers or the price of food at a health facility can be insuperable obstacles. Patients residing in urban centers usually have a choice of alternative providers and treatment options, referred to in the model as *public treatment* or *private treatment*. Typically the distinguishing features of private facilities are higher prices and greater confidentiality. Patients in rural areas may have available only one facility in each class, and sometimes the closest private facility will be quite far away. A third option, *augmented public treatment*, consists of standard public treatment with the addition of a government-subsidized nongovernmental organization (NGO) to help patients adhere to the prescribed behaviors and medications.

Figure 4.2 Choice of Treatment Mode among Symptomatic HIV-Positive Adults



Source: Authors.

This third option has existed in Thailand for more than a decade because of the initiative of international NGOs such as Médecins sans Frontières (MSF). The concept of the grant-subsidized support group, typically consisting of persons living with HIV/AIDS (PHAs), has recently received endorsement from the government and financial support from the Global Fund to Fight AIDS, Tuberculosis, and Malaria. The number of such groups is thus expanding.

To move from the conceptual model of figure 4.2 to a simulation model capable of plausibly describing and predicting real-world behavior, we must choose a functional representation of the impact of price, availability, and other variables on the patient's choice of treatment mode. To avoid the complexity of feedback from the epidemiological model to the demand model, we will assume that mode characteristics affect the proportion of patients choosing each mode without regard to the number of such patients.[2] Furthermore, we seek a functional form that guarantees that under any combination of mode characteristics the proportions of patients choosing the four treatment modes adds to unity (to 1.0). The functional form most frequently adopted in the literature is that of the multinomial logit, which has a respected history of modeling consumers' choices among multiple alternatives partly because it has these two desirable characteristics.[3] In our application, the demand for public ART using this functional form is written as follows:

$$q\_pub\_art = \frac{e^{(\alpha\_pub + \beta \cdot LN(Ppub\_art) + \gamma \cdot LN(Dpub\_art))}}{\sum_{e}(\alpha\_x\_art + \beta \cdot \mathrm{LN}(Px\_art) + \gamma \cdot LN(Dx\_art))}$$

$$x = \{\ pub, apub, priv, no\}$$

where

LN denotes the natural logarithm

e denotes the base of the natural logarithms

q_pub_art = the proportion of patients choosing public ART

Px_art = the price paid by patients for delivery mode *x*, defined as one of the following:

pub = public mode of ART delivery

apub = augmented public mode of ART delivery

priv = private mode of ART delivery

no = treatment for opportunistic infections without ART

Similarly:

Dx_art = the average distance from patients to delivery mode *x*, as defined above

α*x_art* = a constant that captures all other features of mode *x* that influence the patient's choice, such as quality and information

β = the elasticity of patients' demand to price

γ = the elasticity of patients' demand to distance.

The demand functions for the other three modes look the same except that the mode designation *pub* in the dependent variable and in the numerator is replaced by the appropriate mode designation.[4]

The variables *Ppub_art*, *Dpub_art*, *Papub_art*, *Dapub_art*, *Pno_art*, and *Dno_art* are all policy instruments that the government can manipulate to affect patient mode choice. Hence, our alternative scenarios consist of alternative projections of these variables. To predict the pattern over time of the share of patients across modes, one must choose values of the β and γ parameters and of the four values for α*_x*. We have calibrated these parameters for two groups of patients: the symptomatic and the asymptomatic. Table 4.1 gives the two elasticities and four intercept values for each patient category.

The elasticity parameters can be interpreted as the percentage change in the odds ratio between two mode choices in response to a 1 percent change in the ratio of their prices or distances.[5] Therefore, the elasticity of –0.5 of symptomatic people to the price of ART means that the odds of a person choosing any of the modes over another falls by 0.5 percent for every 1 percent increase in the price ratio between the modes.[6] Because sick people have difficulty travel-

Table 4.1 Calibrated Values of the Parameters That Characterize Patient Choice of Treatment Mode in Thailand

	Patient category	
Parameter	***Symptomatic***	***Asymptomatic***
Price elasticities	–0.50	–1.00
Distance elasticities	–1.50	–0.50
Intercepts		
Public ART	7.08	4.42
Augmented public ART	8.14	3.52
Private ART	10.15	4.49
No ART	1.00	1.00

Source: Authors' calibration of Thai historical aggregated data for 2001–3.

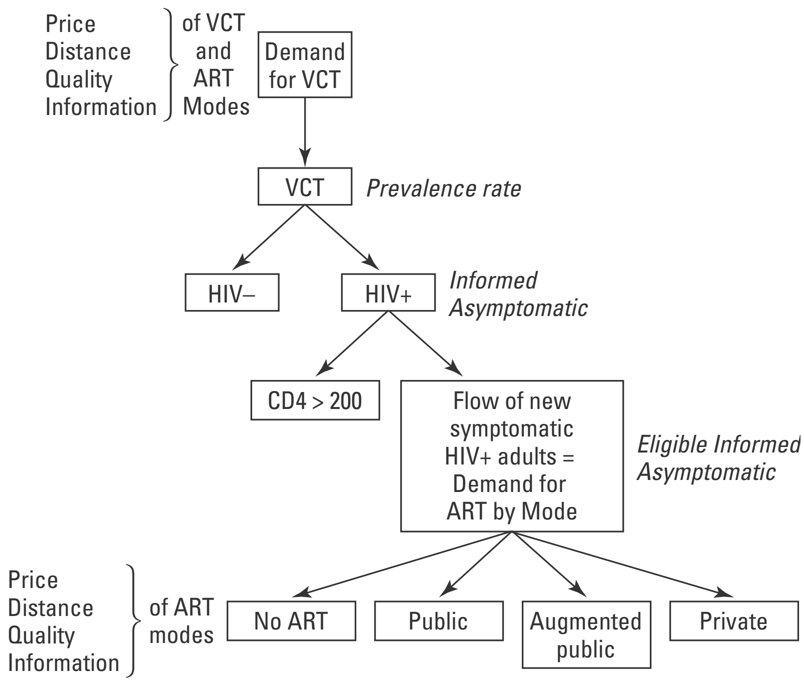
ing, we assume that they will be quite responsive to distance, with a calibrated elasticity of –1.5. The intercept parameters are normalized at a value of 1 for no ART. The ratios of the other intercepts to the "no ART" intercept reflect the perceived desirability or subjective quality of care in the patients' eyes. Given the effectiveness of ART in relieving symptoms and prolonging life, it is not surprising that the intercept parameters for all three treatment options are much larger than for the "no ART" option. Among the three treatment options, private care seems to rank highest, perhaps because of the greater confidentiality as well as the personalized service. Augmented public is next largest while pure public is somewhat smaller, perhaps reflecting a perception that augmentation by an NGO improves the effectiveness of care.

Figure 4.2 depicts our conceptual model of the decision to seek treatment made by a person who is suffering symptoms of AIDS (symptomatic HIV-positive) and visits health care facilities to seek alleviation of that suffering. Most patients who seek treatment in Thailand fall in this category and have advanced AIDS disease with CD4 counts of 50 cells per cubic millimeter or below. However, current medical opinion holds that the optimal time to begin ART is significantly earlier, when the patient's CD4 count has dropped from its normal level of between 600 and 1,000 to a level of 200 cells per cubic millimeters. Most such patients are still symptom-free and would be unaware of the benefit they could obtain from ART, unless they had taken a test for HIV infection. The path to treatment for these asymptomatic patients thus begins with the demand for VCT.

Figure 4.3 illustrates our conceptual model of this route into treatment. The individual must first decide to seek out the VCT service, a decision that depends on the characteristics of those services and on the patient's information about them. Once the individual learns that he or she is HIV positive, a count of the density of CD4 cells in his or her blood determines whether treatment should be started. Asymptomatic patients who demand care have passed through these two filters.

We model the first filter (VCT) using a demand curve like that for ART but simplified to include only two options: "seek VCT" and "do not seek VCT." Again, the policy instruments available to government include a price and a distance, the price of VCT and the average distance to it. But in addition to these two instruments, which act directly on the demand for VCT, we posit that patients demand more VCT if

Figure 4.3 Demand for Voluntary Counseling and Testing and Choice of Treatment Mode among Asymptomatic HIV-Positive Adults



Source: Authors.

the price of and the distance to public ART are lower. This assumption captures the frequently suggested hypothesis that people are more willing to learn their HIV status if they know that treatment is available to those who test positive.[7] Accordingly, in our model, policies followed by the government to encourage ART also encourage VCT.

The demand for VCT depends not only on characteristics of its supply but also on characteristics of prospective candidates. People who believe that they have been exposed to HIV through risky behavior have a greater incentive to seek information about their serostatus, especially when ART is available. The reverse is true among low-risk people, who may be hard to convince to seek VCT. To capture the difference between these two populations of consumers, we specify and calibrate separate demand curves for high-risk and low-risk groups.

The calibrated values of the parameters of the VCT demand function are given in table 4.2. The selected values of these parameters reflect our belief that the high-risk group is highly responsive to the availability of ART—more so than the low-risk group, which does not perceive itself to be in danger. In contrast, the low-risk group is highly responsive to the price of VCT but less likely to consider the price of

Table 4.2 Calibrated Values of the Parameters That Characterize Patient Choice of VCT

	Risk group	
Parameters	***Low risk***	***High risk***
Price elasticities		
Price elasticity of VCT	–1.5	–0.2
Price elasticity of ART	–0.4	–0.6
Distance elasticities		
Distance elasticity of VCT	–1.5	–0.2
Distance elasticity of ART	–0.4	–0.6
Intercepts	–12.2	–2.2

Source: Authors' calibration of Thai historical aggregated data for 2001–3.

ART in making the decision. For lack of information, we assume the same elasticities for both prices and distances. The values of the intercepts reflect our belief that the high-risk group is much more likely to be tested than the low-risk group.

Once VCT candidates receive their results, they learn whether they are HIV positive or negative. Those who are positive learn whether they have CD4 counts lower than 200 cells per cubic millimeter and thus are eligible for ART. The demand function for these asymptomatic people is given by the parameters in the second data column of table 4.1.

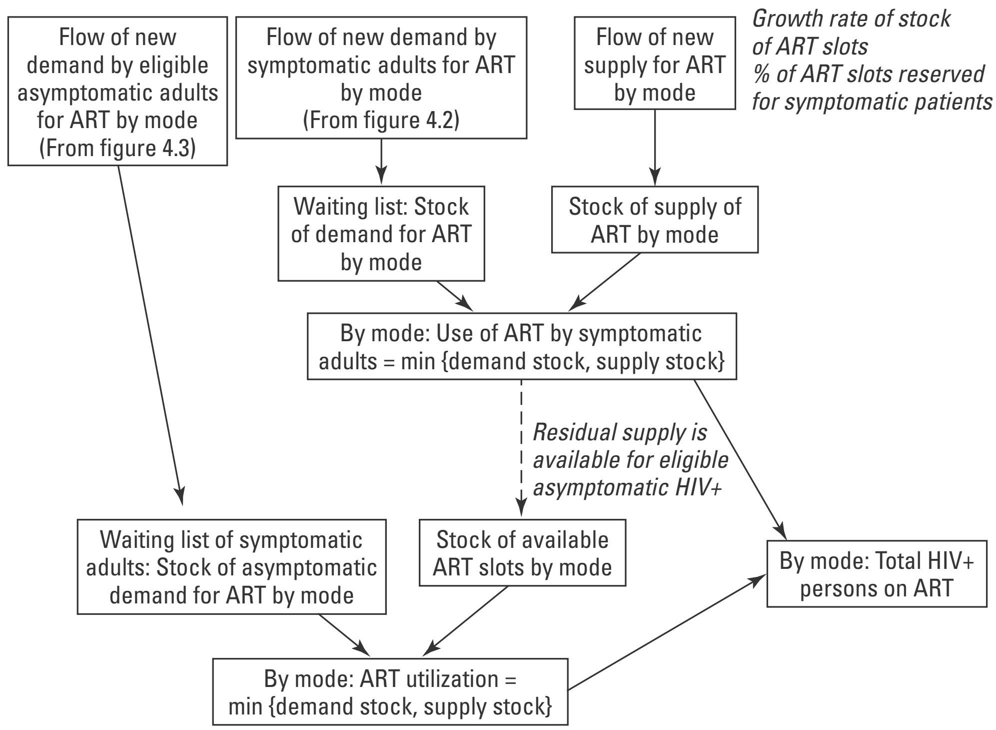
When the supply of ART is less than the demand for it, shortages and rationing will result. Market-clearing mechanisms in such circumstances might include price increases (legal and illegal), longer queuing times, or quality reductions. In the case of acute illness episodes of other diseases—such as food poisoning, acute upper respiratory infections, or malaria—demand is brief, lasting only a few days per episode. Supply must occur simultaneously with demand for an equally brief period. For such diseases, both supply and demand can be modeled as flows of episodes per year. However, once a person starts ART, that person must—with only minor exceptions—remain on ART for the rest of his or her life. In this respect, ART resembles kidney dialysis for those with kidney failure and insulin therapy for diabetics. For all these diseases, the flow of new demand for care in a given year is additional to the stock of demand remaining in the system from previous years. To satisfy all this demand, the stock of available ART treatment slots must equal or exceed the stock of patients demanding ART.

The models we have presented for the demand for ART and for the demand for VCT are models of flows. The demand for VCT is modeled as a proportion of all HIV-negative adults.[8] The demand for ART by asymptomatic people is modeled as a proportion of those who know that they are HIV positive and learn in a given year that they have a CD4 count less than 200 cells per cubic millimeter and are thus eligible for ART. The demand for ART by symptomatic people is modeled as a proportion of those whose CD4 counts fall for the first time below a critical level of 50 cells per cubic millimeter, a level that we assume is on average associated with the onset of AIDS symptoms. Because some people develop symptoms that lead to AIDS treatment before their CD4 counts drop this low, this assumption underestimates the total demand by symptomatic patients and therefore leads to underestimates of government expenditure.

Our assumption about the supply of VCT services is simply that a sufficient flow will be available in unlimited quantities to meet demand at a fixed unit cost. For ART, however, we must take into account that supply is a stock of treatment slots and demand for those slots is expressed by a stock of eligible patients, including some who have been on ART for some time and others who are newly eligible or have been eligible for a while but unable to find a treatment slot. Figure 4.4 illustrates how we compute the actual amount of ART delivered by comparing the stocks of ART supply and ART demand.

The procedure for allocating treatment slots is guided by a policy rule that derives from the conventional medical rule to tend first to those in direst need. We assume that patients who were under treatment the previous year have first priority for treatment slots unless they are lost to follow-up or die. Available treatment slots are then assigned in order of priority so that a specified percentage (*X*) of slots go to the new flow of demand by symptomatic patients and the remaining percentage of slots are available to the new flow of demand from eligible asymptomatic patients. If the government were able to reassign available treatment slots instantly around the country to meet year-to-year fluctuations of demand by newly symptomatic patients, then *X* might be set to 100 percent. In this case, the model would describe a situation in which all demand by symptomatic patients anywhere in the country would be satisfied before the first asymptomatic patient is treated anywhere by the public sector. However, we set *X* to 90 percent to capture the difficulty of achieving such instantaneous reassignment or an inten-

Figure 4.4 Supply and Demand for ART by Treatment Mode



Source: Authors.

tional government policy of treating some people when their CD4 counts are above 100 cells per cubic millimeter. This assumption that 90 percent of slots go first to symptomatic patients is maintained across all the simulated policy scenarios both here and in chapter 5.

In summary, the policy part of our model takes as inputs the policy instruments listed in table 4.3 and provides as outputs to the epidemiological model the variables listed in table 4.4. The input variables subject to government control are primarily prices and distances to testing and treatment facilities. The output variables are the proportions of various population groups that seek VCT or ART in a given year and the availability of ART treatment slots for all the accumulated people demanding care.

The assumptions about the functional form of the demand equations and the calibrated values of the elasticities and other demand parameters are based on the limited data available. In the future, it may be possible to collect microeconomic data on households that are actually making the choices modeled here. Then these data could be used to estimate the values of these demand parameters and to choose the best-fitting functional forms for the demand functions. Since such data are

Table 4.3 Major Policy Inputs to the Policy Model That Generate Alternative Scenarios: 18 Projected Variables

	HIV-positive population		*HIV-negative population (CD4 < 200 cells/mm^3)*	
	Low risk	*High risk*	*Informed asymptomatic*	*Symptomatic*
Demand for VCT	Price and distance to VCT (2 variables)	Price and distance to VCT (same 2 variables)	n.a.	n.a.
Demand for ART	n.a.	n.a.	Price and distance to ART by treatment mode (2 variables)	Price and distance to ART by treatment mode (same 2 variables)
Supply of VCT	Unlimited	Unlimited	n.a.	n.a.
Supply of ART	n.a.	n.a.	Growth rate of facilities (average slots per facility), by treatment mode	Growth rate of facilities (average slots per facility), by treatment mode

Source: Authors.
Note: n.a. = not applicable.

not currently available, we impose more theoretical structure on the problem, to provide plausible projections of the alternative scenarios.

We model government policy options as affecting the future performance of ART delivery through government-mandated changes in the prices of public and augmented public ART, in the provision of public ART, and in the provision and financing of augmented public ART. These mandated changes directly affect demand through the price and distance variables, as described above. They also affect supply (for example, the total available treatment slots). The policy workbook part of our model captures these processes by generating a time series of input values for the epidemiological model. The 18 time series variables generated are shown in table 4.4.

An Overview of the AEM

The Model

The AEM is a difference equation model that projects the dynamic patterns of HIV epidemics in Asian settings. Developed by Brown and Peerapatanapokin of the East-West Center, in collaboration with the

Table 4.4 Outputs from the Policy Model That Drive the AEM for Each Scenario: 18 Projected Variables

	HIV-positive population		***HIV-negative population (CD4 < 200 cells/mm³)***	
	Low risk	***High risk***	***Informed asymptomatic***	***Symptomatic***
Demand for VCT	Proportion demanding VCT (1 variable)	Proportion demanding VCT (1 variable)	n.a.	n.a.
Demand for ART	n.a.	n.a.	Proportion demanding each mode (4 variables)	Proportion demanding each mode (4 variables)
Supply of VCT	Unlimited	Unlimited	n.a.	n.a.
Supply of ART	n.a.	n.a.	Minimum slots for each treatment mode (4 variables)	Maximum slots for each treatment mode (4 variables)

Source: Authors' construction.
Note: n.a. = not applicable.

Thai Working Group on HIV/AIDS Projection and the Ministry of Public Health (MOPH), the model is sufficiently disaggregated to benefit from available data on risk behavior and HIV prevalence in all the important risk groups in the Thai population (Brown and Peerapatanapokin 2004b; Thai Working Group on HIV/AIDS Projection 2001). Before the addition of modules to capture ART, versions of the AEM were presented to and reviewed by attendees at several international conferences and at meetings of the Joint United Nations Programme on HIV/AIDS (UNAIDS) Reference Group on Epidemiology. The biological parameters of the model have been calibrated so that the model's projected prevalence rates match observed rates by risk group.[9] Such detailed fitting is rare among models of HIV epidemics, partly because most countries outside Thailand have too few data to permit such comparisons.

The model has three major transmission modules: heterosexual contact, needle sharing, and homosexual male-to-male contact. These components are subdivided into eight populations:

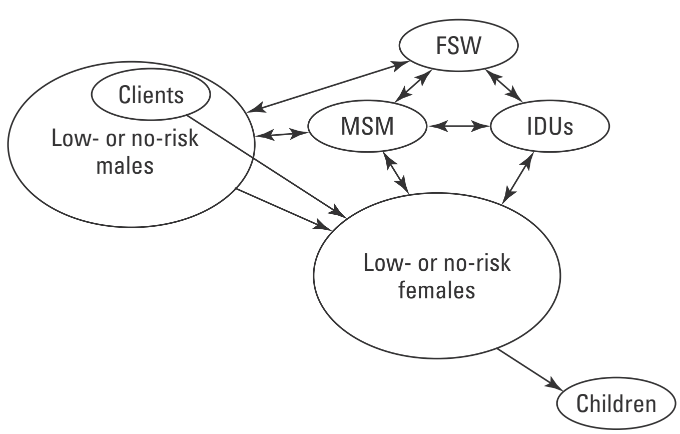
- male clients of sex workers
- male nonclients of sex workers
- male injecting drug users (IDUs) with high needle sharing

- male IDUs with low needle sharing
- men who have sex with men
- female direct sex workers
- female indirect sex workers
- female nonsex workers (that is, the general female population).

Interactions of these subpopulations are illustrated in figure 4.5.

Each of these subpopulations is further divided into those who are infected (HIV positive) and those who are susceptible to infection (HIV negative). The AEM calculates HIV infection in the seronegative members of each risk group as driven by that group's risky contacts with the other groups and the HIV seroprevalence in each of those other groups. The model also makes assumptions about transitions from one risk group to another.[10] These assumptions allow the model to capture the common observation that individuals are not permanently identifiable as members of a specific risk group but change their behavior over their life cycles. Heterosexual men are likely to visit sex workers more frequently when they are single than when they are married; many female sex workers practice their trade for only a few years before retiring to get married; men who have sex with men—and, indeed, all adults—have fewer partners when they get older (Brown and Peerapatanapokin 2004b; Guest, du Guerny, and Hsu 2003).

Figure 4.5 Stylized Structure of the AEM: Transmission



Source: Brown 2004.
Note: FSW = female sex workers.
MSM = men who have sex with men.

New infections in any given group are hence a function of eight quantitative factors, all but one of which changes over time. For example, consider the equation for the rate of new infections among the uninfected male clients of female sex workers. This infection rate is affected by numerous factors:[11]

- the prevalence of HIV among sex workers
- the fraction of sex acts that are not protected by condoms
- the frequency of contact (that is, the frequency of the client's willingness to pay for sex with a sex worker)
- the proportion of clients who are circumcised
- the reduction in the probability of infection attributable to circumcision
- the proportion of clients who have a classic sexually transmitted infection (STI) such as syphilis or gonorrhea
- the degree to which such an STI amplifies transmission probability.

The one constant parameter in the model, the biological rate of transmission between an infected and an uninfected person in the two contact groups, is calibrated so that the projected incidence, prevalence, and behavioral patterns match the observed historical time paths for these variables.

Among the most important behavioral input variables included in the model are

- sizes of key populations over time—for example, total population, number of 15-year-olds in each year, IDUs, men who have sex with men, and direct and indirect sex workers
- frequency of risk behaviors, including sex with various partner types, sharing of needles, and so on
- levels of protective behaviors—for example, use of condoms among different risk groups, use of clean needles, and percentage of men who are circumcised (Thai Working Group on HIV/AIDS Projection 2001).

Asian Epidemic Model with ART

Until 2002, the number of patients on ART in Thailand was too small to have a significant effect on the prevalence rate of HIV infection or on the rate of new infections (incidence). As a result of the Thai government's decision to adopt the NAPHA policy, both the numbers of people receiving treatment and the presumed effectiveness of treatment have dramatically improved. Accordingly, it is no longer possible to make reasonable projections about the course of the epidemic without including the effects of ART. The authors of the AEM joined the report team to model as accurately as possible the potential effects of ART policy on ART use and the effects of ART implementation on the course of the epidemic. This collaboration came about for two reasons:

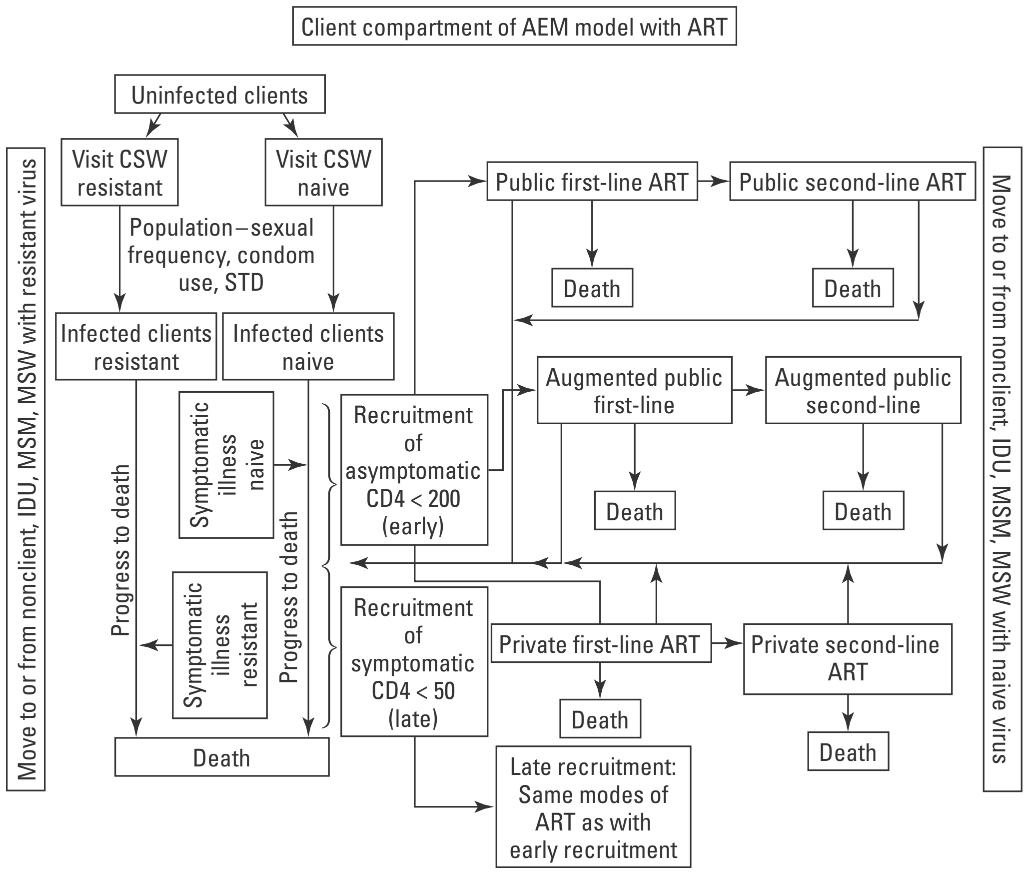
- to update the AEM to current and future epidemiological realities
- to respond to the Thai government's request to model the costs and effects of their new treatment policies and its possible variants.

It has led to the development of a new "policy workbook" model to capture the effect of policy on ART use and to a revised version of the AEM that integrates into the model the complexities of the Thai system for providing ART.

The ART components of the revised AEM build on earlier work to integrate ART into a simpler differential equation model of the AIDS epidemic in India (Over and others 2004). However, the ART components of the AEM differ from that earlier work in important ways that reflect Thailand's greater experience with public sector ART provision, its ability to afford a more generous ART treatment protocol, recognition of the challenges that all countries confront in recruiting patients before their disease advances to the point that they are difficult to help, and the more sophisticated modeling foundation of the AEM.

Figure 4.6 uses the ART component of the model for male clients of sex workers as an example of how ART processes are appended to the epidemiological superstructure of the AEM. The left side of the figure is a simple representation of the epidemiological part of the model for clients. The first new feature introduced to the model because of ART is the possibility that the client will be infected with a strain of HIV that is resistant to ART. Such strains did not exist at significant levels in Thailand until the advent of the new, ambitious treatment policy. One reason for adding ART to the AEM is to project the effect of various ART policies on the spread of such resistant strains. In the present implementation of the AEM, we assume for simplicity that people

Figure 4.6 Client Compartment of the AEM with ART



Source: Authors.
Note: MSM = men who have sex with men, MSW = male sex workers, CSW = commercial sex workers.

infected with a resistant strain do not benefit from ART and therefore do not use it. As Thailand continues to gather information on the prevalence of resistant strains and as the public sector develops and widely implements therapeutic protocols for such patients, the AEM can be modified to include those additional complexities.

The numerically more important branch on the right side of figure 4.6 is infection with a naive (or "wild" or "susceptible") strain of HIV. The progression of disease in these patients is modeled as an inexorable process according to the natural history assumptions detailed in chapter 3, unless it is interrupted by treatment. The HIV-infected person has two opportunities to be recruited for treatment. If the person seeks VCT and learns that he (or in other compartments of the model, she) is infected and has a CD4 count below 200 cells per cubic millimeter, that person might choose to seek treatment relatively early during the illness (from two to six years after infection). The AEM captures here the outcome of the choice process that is modeled in the policy workbook model and described above.

Once the person is recruited into ART and selects one of the three treatment modes, he or she follows the progression to treatment failure or to dropping out of treatment that is characteristic of that treatment mode and recruitment time. The figure displays the possibilities of remaining in first-line therapy, dropping out of it, moving on to second-line therapy, or dying. From second-line therapy, the figure displays the possibilities of remaining in therapy, dropping out, or dying. Those progression probabilities are modeled as described in chapter 3 and are based on all available data and on the expert opinions of the clinicians on the report team and among the background paper authors.

If the person does not learn his or her HIV status or, despite learning that he or she is positive and eligible, chooses not to seek early treatment, the person can again be recruited for treatment when he or she becomes symptomatic at CD4 counts below 50 cells per cubic millimeter. The decision to seek treatment at this late stage is again modeled in the policy workbook. The outcome of individual decisions is captured here in the AEM by the box labeled "late recruitment." The structure of treatment options available to patients at late recruitment is identical to those available at early recruitment.[12] However, as described above, progression to treatment failure and disease is more rapid for the average patient than it would have been if he or she had begun treatment with a more intact immune system. Also, the public health care system is presumed to give priority to symptomatic patients if it is ever short of treatment slots. We assume that only 10 percent of treatment slots are reserved for early treatment regardless of the size of the asymptomatic waiting list.[13]

Measuring the Performance of ART Policy

The objective of ART policy is to improve the health of patients, allowing them to live longer, healthier lives. But ART policy has other effects—some beneficial to the patient and society and some potentially harmful. A full economic analysis of ART would attach a monetary value to all of its effects, whether beneficial or harmful, to compute the net present value of the policy. Such an aggregated perspective is especially useful for comparing the social rate of return on investments in treating AIDS patients with the return on investments in other sectors. The disadvantage of adopting the full economic approach lies in the fact that unit values for healthy life-years, years of orphanhood, and other effects of the ART policy are hard to establish or defend. Analysis based on controversial unit values is itself controversial, and this controversy may distract from points that can be made without adopting these monetary values. So in this analysis we have chosen to keep track of as many effects of ART as possible, without attempting to aggregate them across categories. Our approach fits within the framework of cost-effectiveness analysis rather than cost-benefit analysis.

Effects

Before the advent of ART in developing countries, the best indicator of the spread of HIV infection was the prevalence of infection in the population. Although economists and epidemiologists have complained that the best measure of the effectiveness of AIDS prevention policy would be the change in the incidence rate (that is, the rate of new infections), data on the incidence of HIV is so rare that prevalence was used in its place (World Bank 1999). Typically, a decrease in the prevalence rate was interpreted to mean that the incidence rate must be smaller than the rate of death of HIV-infected people. This was the pattern observed in Thailand in the 1990s (World Bank 2000). In the absence of ART, an increase in prevalence could be interpreted to mean only that the incidence was higher than the death rate. This situation could occur in a country such as South Africa, where infection spread rapidly before the death rate caught up, or in a country with a longer history of widespread infection because incidence continues to increase. In either case, in the absence of effective AIDS treatment, rising prevalence would always be bad news.

The presence of ART has reversed the interpretation that must be placed on increasing prevalence levels. By keeping infected people alive longer, ART maintains a prevalence rate even if incidence falls to zero. With a small incidence rate and an effective ART program, prevalence will continue to rise. Prevalence retains its value as an index of the current and potential future burden of the epidemic on the country's health care system, but it loses its value as an indicator of the effectiveness of an HIV prevention policy.

The number of deaths is an index of the effectiveness of an ART program that is particularly useful in the short run. In the years immediately following the introduction of ART access, the number of annual deaths falls almost one for one with the number of treatment slots made available. However, deaths are not a valid index of the continuing benefits of an ART program over the long run. Because everyone dies eventually, comparing a future without ART to one with ART shows the effect of ART to be fewer deaths in the near future and then even more deaths in the more distant future. This effect occurs because the cohort of people placed on ART at the beginning of the program begin to die, which adds their deaths to those of the HIV-negative people whose dates of death are not affected by the ART program. It will appear that the ART program is causing deaths a decade after its introduction, when it is actually having the beneficial effect of shifting the deaths forward in time. Hence, the number of deaths averted, like HIV prevalence, is an unappealing measure of the effectiveness of an AIDS treatment policy.

Because the objective of ART policy is to lengthen and improve the lives of the recipients, a natural measure of the effectiveness of a policy is the number of life-years it adds to the population. With a microsimulation model, such a measure is easy to implement by simply keeping track of the number of life-years lived by each individual and comparing the aggregate of these life-years between two runs of the model: one with ART and one without it. Because the AEM is a difference equation model, in which individuals are subsumed into groups, it is not possible to directly measure the life-years saved. However the model does provide a count of the number of people beginning ART at each of the two recruitment points: early and late. Figure 3.4 in chapter 3 presented a calculation of the estimated number of additional life-years that accrue to each individual who begins treatment, sorted by whether that treatment is early or late. We measure the effectiveness of an ART policy option by multiplying this estimate of per patient life-year benefit by the number of patients initiating each type of treatment in that year. This method gives a stream of annual benefits that continues over time as long as treatment continues.

Another potential benefit of ART is that children are not orphaned until they are older and farther along in their transition to independence. That this benefit is potentially important, not only for the psychological well-being of the children directly concerned but also for national economic growth, has been demonstrated in a recent series of papers by Bell, Devarajan, and Gersback (2004). However, the AEM is not a full demographic model and therefore does not capture fertility in a realistic way. The report team has not attempted to estimate the years of orphanhood that would be saved by ART treatment.[14]

Chapter 3 described the possible effects of ART on preventing new HIV infections. Through the biological channel, ART might either slow transmission (by reducing the viral load in bodily fluids) or accelerate it (through greater or longer sexual activity of the infected person). Through the behavioral channel, ART could also have either beneficial effects (by encouraging the demand for VCT) or negative effects (if people relax their efforts to prevent infection in response to better availability of high-quality treatment). Furthermore, the degree of adherence practiced by the average ART patient also affects the spread of resistant strains of the virus—a potentially dangerous negative spillover effect of treatment on others. The index of all of these effects is the rate of new HIV infections or the HIV incidence rate.

In principle, it is desirable to measure the effects of ART policy on the distribution of the burden of HIV/AIDS by income category. It is often asserted (with little empirical support) that the Thai HIV epidemic mainly affects the poor. If this assertion is true, a generous ART subsidy might be equity enhancing, helping the poor more than the rich. However, the AEM does not distinguish groups of people by socioeconomic status. Accordingly, it affords little opportunity for analyzing the equity effects of alternative ART policies. However, chapter 5 considers some equity implications of NAPHA based on population averages.

In comparing alternative AIDS policies, our analytical techniques must recognize that the effects of current policies persist for a long time. When the Thai government buys vehicles, it can hope that they last five years. Roads last 5 or 10 years, and buildings last 15 or more years. However, the consequences of AIDS treatment policy will last even longer, as cohorts of people remain alive rather than dying, infect or don't infect others, and continue to consume medical care for as long as three decades after beginning treatment. Therefore, it is essential to adopt a long-term perspective in measuring the effects of alternative polices. To consistently compare scenarios in which the benefits and costs occur at different points in time, this report applies standard discounting formulas to future streams of benefits and costs to convert them to their present values (World Bank 1993).[15]

Costs

Cost analysis of health programs typically distinguishes between fixed costs and variable costs. The fixed costs are occasioned by the establishment of a health program, while the variable costs are typically unit costs multiplied by the number of units of output. In this simple cost model, the fixed costs are spread over all the units of output, leading to the phenomenon of increasing *economies of scale:* the more units of output produced, the lower the cost per unit of output. When programs are added to existing programs, costs are typically saved because some of the fixed and variable costs of the new program can be avoided by using existing facilities and resources. Economies of this sort are called *economies of scope*. When a program is expanded beyond its most efficient scale or extra programs added to existing ones get in the way of one another, diseconomies of scale or of scope can also appear. The question is whether the Thai ART program will experience economies of scale or scope and, if so, whether those economies will increase or decrease unit costs.

For the purposes of projecting the costs of ART policy, we have adopted the simplifying assumption that most program costs relate to the individual patient—in the form of provider time, pharmaceutical products, diagnostic tests, and disposable paper and rubber products—and therefore do not vary much with scale or scope. The exception we make to this rule is the cost of equipping a health care facility that is not a district hospital with the capacity to administer and manage patients on ART. We assume that to qualify to manage one or more ART patients, any facility must train a minimum number of providers in ART protocols and, to keep them abreast of the rapidly changing technology, must retrain those staff members every year. This category of costs might be called *recurrent fixed costs*. They recur every year as a function of the number of facilities, but they can be spread over all the patients at the facility.

We anticipate that once ART treatment is available at all district hospitals, perhaps by 2008, the government will recognize that expanding coverage requires expanding the number of treatment facilities beyond the district hospital network. One option is to equip a selected number of health centers with a physician and a nurse who are trained in ART. The costs of this additional training and retraining will be repaid with the greater willingness of the population living close to these centers to seek treatment and to adhere to treatment regimens.

Projection Scenarios

The impact of a policy choice can be defined only in comparison with what would have happened in the absence of the choice. This alternative scenario, called a *baseline* or *counterfactual*, is a projection of the future course of AIDS treatment had the Royal Thai government not introduced its expanded NAPHA. This section describes how the report defines the baseline and the NAPHA scenarios.

Baseline and NAPHA Projections

Several alternative baseline scenarios could have been chosen: cells (a), (b) and (c) in table 4.5. Each scenario corresponds to a different combination of public financing of ART and government subsidization of the production and sale of low-cost ART drugs. The chosen baseline corresponds to cell (a): what would have happened had the government kept only its pre-2001 voluntary program, with only branded drugs available.

The effect of NAPHA is obtained by comparing outcomes from cell (a) with those from cell (d). The total impact can be separated into the part caused by the availability of low-cost generic ART and the part caused by the public finance of ART provision. Such a deconstruction enables us to attribute portions of the benefits of NAPHA to each of its two components and a portion to synergy between them. We do not undertake that deconstruction here.

Table 4.5 Potential Baseline Scenarios or Counterfactuals to NAPHA

		Government to finance ART publicly	
		No (private out-of-pocket only)	*Yes*
Government to produce and sell low-cost ART (GPO-vir)	*No*	(A) Baseline scenario: No government intervention takes place. There is a voluntary program only, too small to make a difference (for example, no ART).	(B) Government provides subsidized public production with no possibility of alternative supply channels (buyers' clubs and so forth).
	Yes	(C) GPO produces and markets GPO-vir at current prices (less than US$1/day), but government does not expand public delivery of ART through the public health system beyond the voluntary program.	(D) NAPHA: This scenario includes the current form and alternative versions, including stimulating VCT for earlier recruitment and introducing demand-side incentives to increase adherence.

Source: Authors.

Assumptions

Table 4.4 categorized the 18 variables, which are generated by the policy model and then drive the epidemiological model. Table 4.6 presents the specific assumptions for the baseline, which characterizes what would have happened had NAPHA not been implemented, and for the changes made by the government, which define the NAPHA policy.

The Thai government implemented the NAPHA policy by changing a number of policy instruments, which in turn affected people's

Table 4.6 Assumptions for Baseline and NAPHA Scenarios

	Scenario		
Parameter	***Baseline***	***NAPHA***	***Change (%)***
1. Price of VCT (baht)	30	30	0
2. Weight of short-run VCT demand in total (%)	100.0	100.0	0
3. Growth rate of VCT facilities (%)	0.0	0.0	0
Prices of ART per year (baht)			
4. Public	8,530	650	–92.4
5. Augmented public	8,530	1,880	–78.0
6. Private	13,800	9,534	–30.9
7. No ART	30	30	0
Quantities of ART in 2003 (number of facilities)			
8. Public	119	860	622.7
9. Augmented public	45	100	222.2
10. Private	100	100	0
11. No ART (residual)	11,042	10,282	–6.9
Other ART supply parameters			
12. Growth rate of ART facilities (%)	1.0	1.5	50.0
13. Growth rate of augmented ART facilities (%)	1.0	5.0	500.0
14. Growth rate of private ART facilities (%)	1.0	3.0	200.0
15. Starting number of treatment slots per public facility (average)	17.6	54	206.8
16. Starting number of treatment slots per augmented public facility (average)	56	56	0
17. Proportion of treatment capacity designated to symptomatic patients before asymptomatic patients with CD4 < 200 cells/mm^3 are accepted (%)	90.0	90.0	0
18. Growth rate of all public health facilities (%)	1.0	2.0	100.0
19. Number of treatment slots in 2002	2,095	8,341	298.1
20. Number of treatment slots in 2003	2,095	16,663	695.4

Source: MOPH data and authors' estimates.

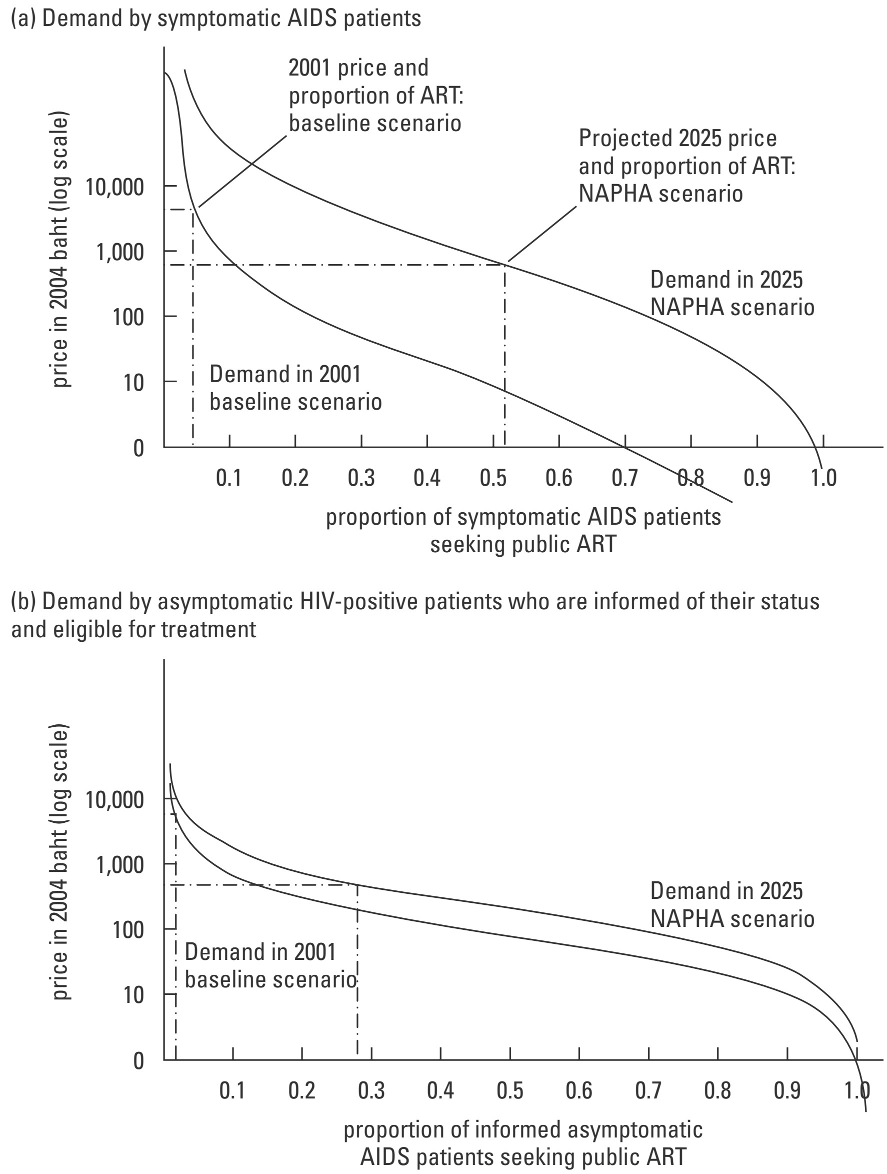
willingness and ability to seek out and use VCT and ART services. Among the most important instruments were the price paid by patients for ART and the distance patients had to travel to obtain it. Table 4.6 shows that the price paid by a patient for ART in a public facility was reduced by 92 percent, and the number of treatment sites was increased more than sixfold, as a result of the NAPHA policy. Because GPO-vir was also available to NGOs and the private sector, the effective prices of treatment fell for those modes of treatment also, though not by as much. We assume that their availability also improved.

Furthermore, in adopting NAPHA, the government launched a process of expansion whose full extent is not yet known. For the purposes of this model, we assume that the number of treatment facilities continues to grow through the end of the projection period by 1.5 percent per year, while the number of patients treated in any given facility expands to meet virtually all available demand. We assume that the NAPHA policy is friendly to but not aggressively promoting the incorporation of PHA groups into the public treatment protocols, so that augmented public treatment grows at 5 percent a year. We assume that private treatment facilities also become increasingly available, with a growth rate of 3 percent a year, though at a substantially higher price.[16]

This model captures our assumptions about the effects of these dramatic policy changes on treatment uptake. Panel (a) of figure 4.7 illustrates the effects on the demand for ART by symptomatic HIV-positive people of the price reduction and the improved availability. Under the 2001 policy, with the price of care at B 8,530 and only 119 facilities delivering care, the proportion of symptomatic AIDS patients nationwide who used care effectively was about 2.4 percent. Another 1 percent of those needing care were being treated in NGOs (that is, the augmented public sector), and we estimate that about 26 percent were being treated in the private sector.[17]

As a result of the reduction in price to B 650 and the increase in availability projected to occur over the next two decades, use will increase for all modes. The second demand curve in panel (a) of figure 4.7 shows the effect of the changes in availability of public ART, while the effect of lower price is shown by moving down from B 8,530 to B 650 on the vertical axis. To enhance the readability of the figure, the price axis is measured on a logarithmic scale. In the new equilibrium

Figure 4.7 Demand for Public ART as a Function of its Price to the Patient



Source: Authors.

in 2025 that will result from this combination of policies, 49 percent of the symptomatic AIDS patients will use public sector ART. Another 24 percent will use augmented public or private sector treatment, leaving 27 percent too distant, too poor, or too ill informed to take advantage of public treatment.

Panel (a) can be used to analyze the relative contribution of the price decrease and the availability increase on projected improvements in ART use. According to this demand structure, the price decrease alone would increase ART use only from 0.2 percent to slightly more than 1 percent of symptomatic individuals. Improved availability alone would increase ART use from 0.2 percent to about 20 percent of symptomatic people. However, if this structure is approximately correct, further price decreases by the public sector are likely to increase its share of total AIDS patients, partly at the expense of private care, but partly from among those otherwise unable to afford it.

Panel (b) tells a similar story regarding demand for care among people who are asymptomatic but have tested positive for HIV and are eligible for treatment. The structure of our demand model assumes that because these people typically have much less severe symptoms or no symptoms at all, they are much more responsive to the price of ART. Because they are less sick, they are more able to travel to seek care. So we assume that they are less responsive to improved availability. Reflecting those assumptions, the demand curves for the informed, eligible, asymptomatic are flatter than those for the symptomatic (that is, less price sensitive) and are closer together (showing less effect from the improved availability).

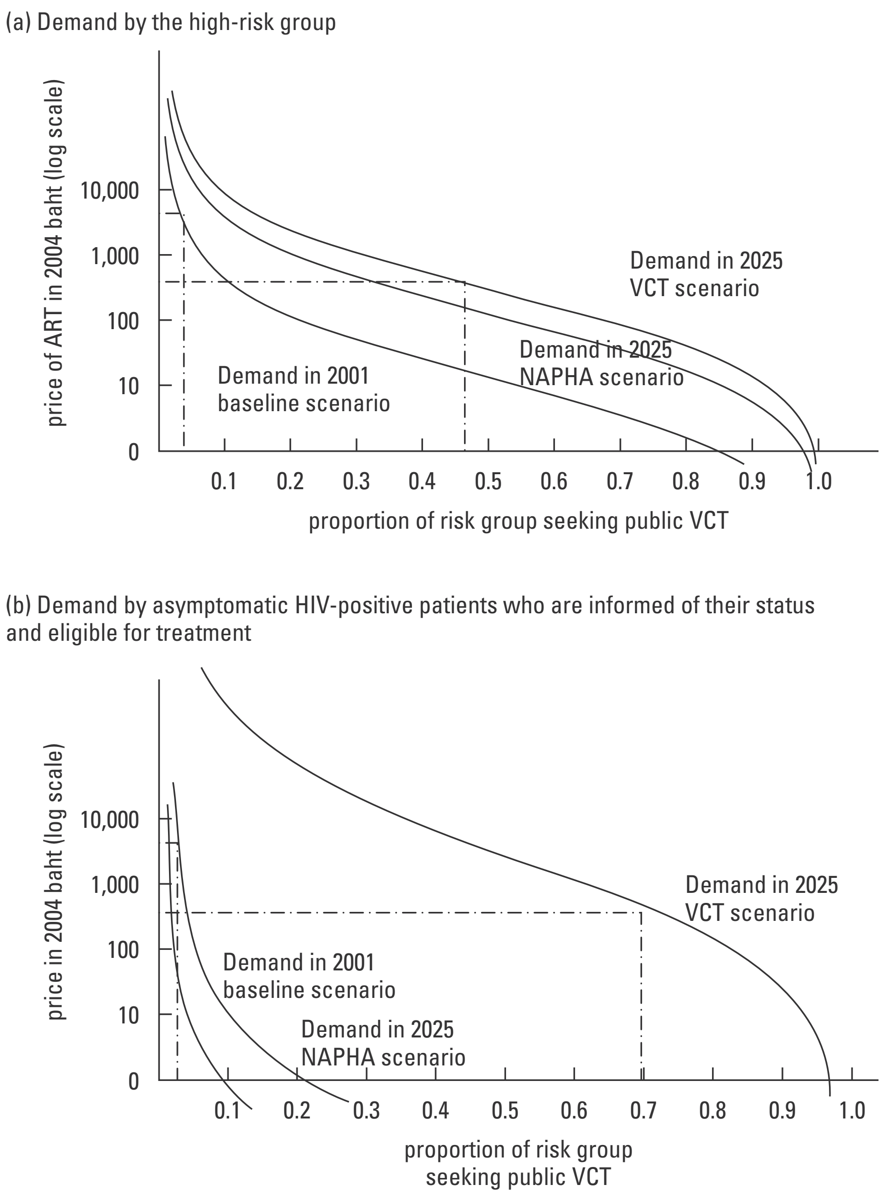
The proportion measured on the horizontal axis of panel (b) in the figure is a fraction of the total number of informed, eligible, asymptomatic, HIV-positive people. According to the structure of the model, for any given point on the asymptomatic demand curve for ART in panel (b), the total number of patients actually recruited depends critically on how many seek testing. Policies that increase the demand for VCT, especially among those at highest risk of infection, increase the number of HIV-positive people who learn of their status and need for treatment early enough to start ART while they are still asymptomatic.

Figure 4.8 presents the demand curves for VCT by the low-risk group and the high-risk group under three scenarios:

- the baseline scenario
- the NAPHA scenario and
- the VCT scenario.[18]

The terms *low-risk* and *high-risk* are composites based on categories of the population in the AEM. Groups from the model that we include in the high-risk aggregate include female sex workers, their clients, men who have sex with men, and IDUs. In the AEM an

Figure 4.8 Demand for VCT as a Function of the Price of Public Sector ART: Baseline, NAPHA, and VCT Scenarios



Source: Authors.

individual is only temporarily in one of these categories and at higher risk of infection before migrating to another, typically lower-risk category.

We assume the price of VCT remains at B 30 in all the scenarios until the end of the projection period. However, we assume that both the price and the availability of ART affect the demand for VCT. Because ART and VCT are complements in consumption, reductions in the price of ART or improvements in its availability stimulate the demand for VCT. Accordingly, it is useful to construct a demand curve for VCT as a function of the price of ART.

Panel (a) of figure 4.8 shows the effect on the proportion of the high-risk group that demands VCT as a function of the price and availability of ART. The leftmost curve presents the demand structure for 2001, when ART was expensive and only 119 locations offered it. About 2 percent of high-risk people were tested in that year. The second curve shows the effect of improved ART availability under the NAPHA policy as a rightward shift of the demand curve at all ART prices by 2025. The price reduction of public ART from B 8,530 to B 650 is projected to move the equilibrium demand for VCT down that demand curve to a point where about 30 percent of the high-risk group will seek testing.

Panel (b) of figure 4.8 shows the quite different structure of demand for VCT assumed for the low-risk group; it is about 20 times larger than for the high-risk group.[19] The leftmost curve again presents the demand under the baseline scenario, when the price of ART is high and its availability limited. Although the predicted proportion seeking VCT in this group was only about 0.02 percent or about one-tenth as large as in the high-risk group, the absolute number of low-risk people tested would have been almost twice as large as the number of high-risk people tested in 2001. Because HIV prevalence in 2001 was slightly less than 1 percent among low-risk adults and more than 4 percent among high-risk adults, the high-risk group would still have more positive tests. For 2001, we estimate about 3,600 positive test results among asymptomatic people, of which about 2,600 were in high-risk groups.

Because low-risk people may not perceive themselves to be in danger of infection, we model their demand for testing as more sensitive to the price and availability of VCT than to the price and availability of ART. The leftmost two demand curves in panel (b) are steep to represent this lack of responsiveness (or inelasticity) with respect to the price of ART. Furthermore, when the NAPHA policy improves ART availability by 2025, the demand curve shifts only a small amount to the right.[20] The total increase in the proportion of low-risk people seeking VCT in the NAPHA scenario is from 0.2 percent to 1.8 percent. Although this increase is almost tenfold, this small demand suggests that without strong and dramatic new policy measures only a very small minority of the general population will learn their HIV status.

The policy model generates the percentage of people in each risk group who will seek VCT and the proportional distribution of asymptomatic and symptomatic patients across the four treatment modes in each year of the progression period. Figure 4.9 presents these three projections for the baseline situation in 2001 and for the NAPHA scenario in 2025.

This section has shown how the policy model links changes in the policy instruments to changes in the distribution of demand for VCT and ART. These proportional distributions are then provided as input to drive the AEM. The next section examines the health and financial consequences of these inputs in the context of the epidemiological and demographic structure of the Thai epidemic.

Effects of Current Policy (NAPHA)

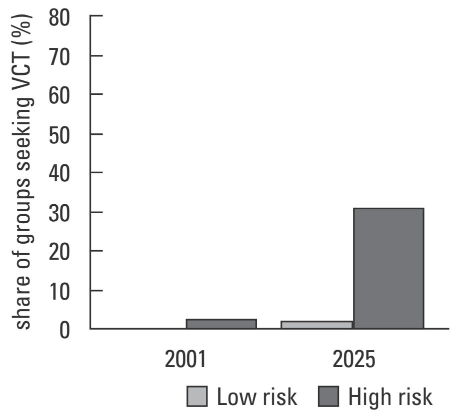
The discussion in the previous section makes clear how dramatic was the change in Thai AIDS treatment policy in 2002.[21] Over just a few months, the price of publicly provided ART dropped by 92 percent, and its availability soared by 620 percent. Because of the new lower price of generically produced ART medications, even the price of private sector ART declined by 30 percent. In this context, projecting the course of the current NAPHA policy is a somewhat arbitrary exercise. We assume that the number of facilities continues to grow at 1.5 percent a year, and we allow the number of persons treated per facility to grow to accommodate all the demand that is elicited by the resulting decrease in the average distance of people from facilities.

Under these assumptions, figure 4.10 demonstrates the result that was predicted in early chapters: by keeping people alive, the NAPHA policy (including second-line therapy) increases the number of HIV-infected people in Thailand compared with what it would have been in the absence of treatment. The model predicts that by 2025 the number of HIV-infected adults will be four times larger than it would have been without NAPHA (263,000 instead of 62,000). Most of this increase occurs because people who would have died of AIDS are living much longer, but some of it is due to increased transmission of HIV by people who are living longer with HIV infection.[22]

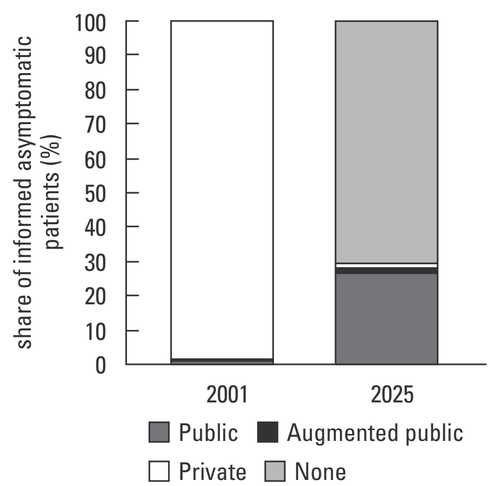
Figure 4.11 shows the effect of NAPHA with second-line therapy on the path of annual total adult deaths in Thailand. A sharp reduction in deaths can be seen at the start of the program—the result of a large cohort of HIV-infected people beginning treatment and surviving

Figure 4.9 Projected Effects of the NAPHA Policy on Demand for VCT and ART

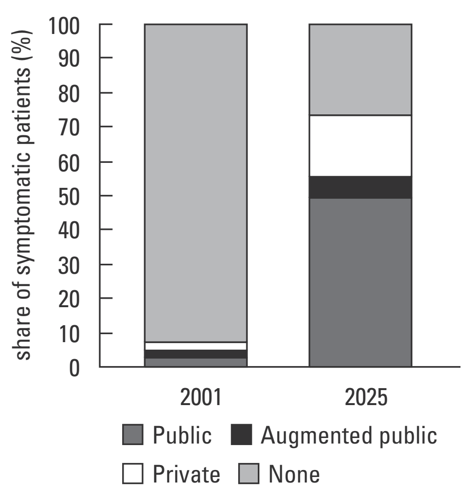
(a) Poportion of low- and high-risk groups seeking VCT in baseline and NAPHA scenarios.



(b) Proportion of asymptomatic patients seeking ART in baseline and NAPHA scenarios.

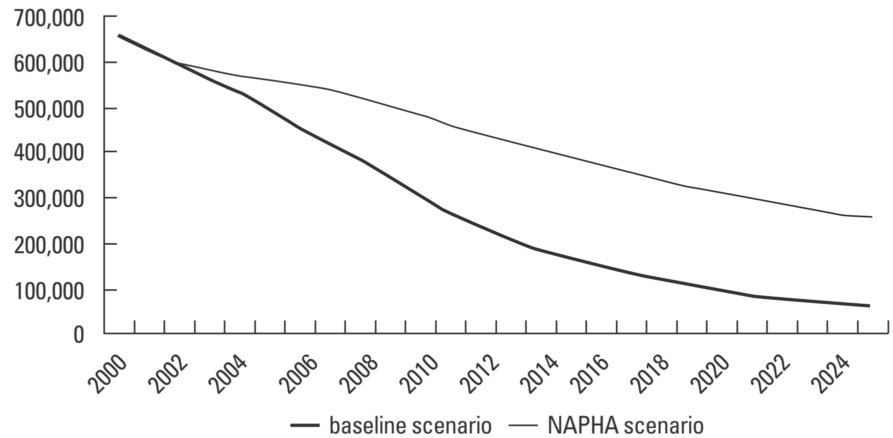


(c) Proportion of symptomatic patients seeking ART in baseline and NAPHA scenarios.



Source: Authors.

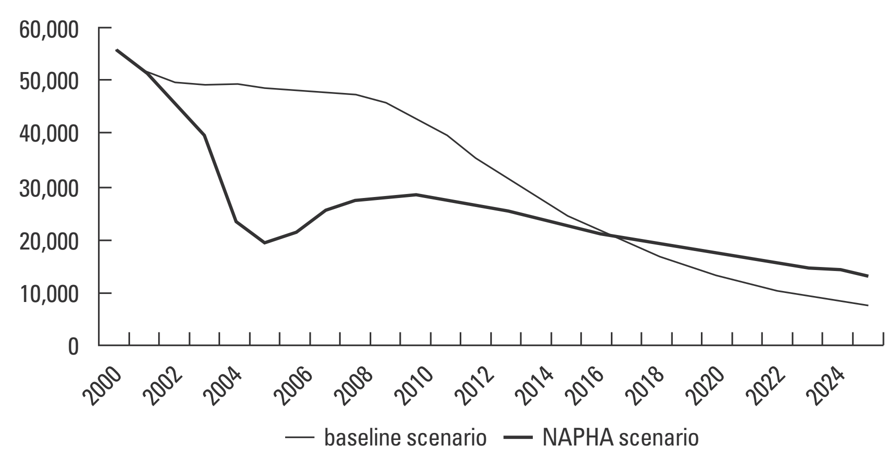
Figure 4.10 Projected Current HIV Cases



Source: Authors.

longer. Figure 4.11 also reveals why "deaths averted" is not a good measure of the success of the ART program. By preventing the deaths of many AIDS patients beginning in 2002, NAPHA postpones those deaths about 12 years. When the patients eventually begin to die, total deaths rebound upward until, in 2017, they exceed the number who would have died that year without NAPHA. The considerable achievement of NAPHA is to shift the survival distribution of AIDS patients to longer survival times, so that the average AIDS patient lives between 5 and 15 years longer. Because everyone dies eventually, the additional deaths after 2017 are not a negative consequence of the program, but a reflection of the longer survival times of the patients.

Figure 4.11 Projected Annual AIDS Deaths

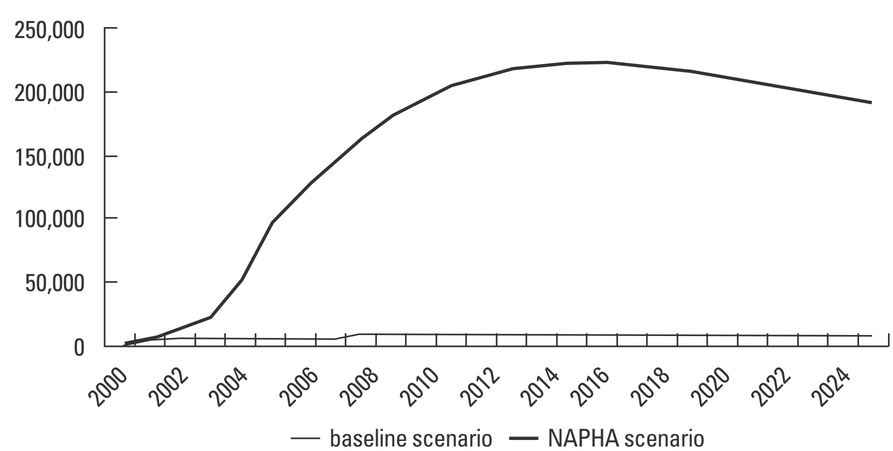


Source: Authors.

The cumulating number of patients under treatment has direct implications for the treatment burden on the health sector. Figure 4.12 shows the extraordinary increase in the national treatment burden attributable to NAPHA with second-line therapy. Under the model's assumptions regarding the benefits of treatment and the declining incidence of new HIV cases, the number of PHAs on treatment under NAPHA peaks in 2015 at about 230,000 and then begins to decrease slowly. Unlike the vast majority of Thai patients, these chronic AIDS patients will have ongoing relationships with the public health system that will last a decade or more. To provide and sustain high-quality care, the system will have to develop and apply modern methods of handling patient records at all the hundreds of treatment sites. Through repeated experience with the health care system, patients will become increasingly informed and thus increasingly likely to point out deficiencies they perceive. So the task of managing communications with this growing patient population will increase in importance from both a medical and a political perspective.

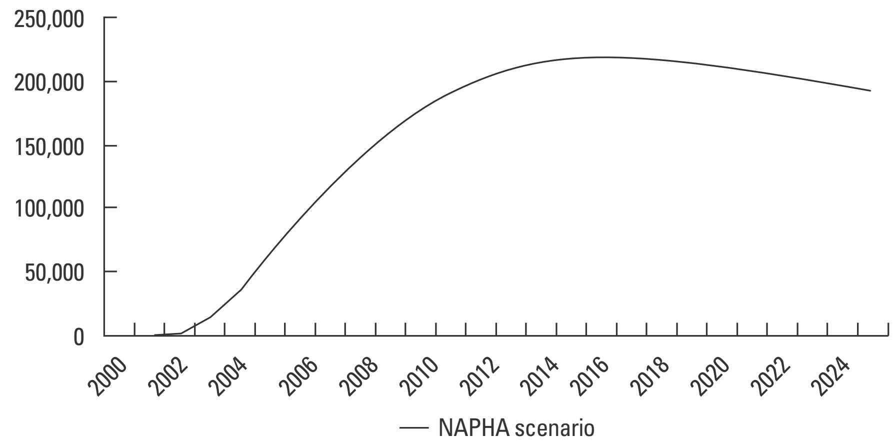
As a measure of the direct benefits of an ART provision policy we use the number of life-years saved, computed as the number of additional people alive in any given year.[23,24] Figure 4.13 presents the predicted flow of saved life-years that result from the implementation of NAPHA with second-line therapy. This figure illustrates the dramatic benefits achieved through the NAPHA policy. By 2015, it will have added about 220,000 people to the living population. Even at the end of the projection horizon, when the Thai AIDS epidemic is predicted to be less severe, the NAPHA policy will save about 190,000 life-years each year.

Figure 4.12 Projected Number of HIV-Positive People on ART



Source: Authors.

Figure 4.13 Projected Annual Flow of Life-Years Saved under NAPHA



Source: Authors.

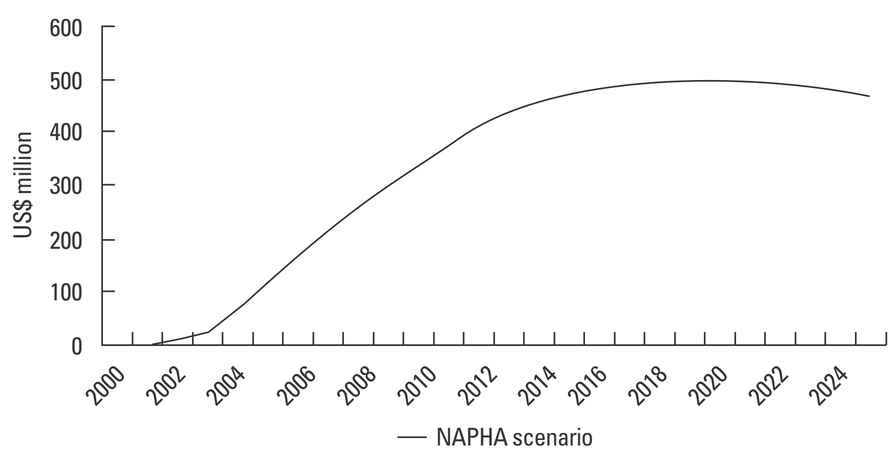
Costs and Cost-Effectiveness of Current Policy

What are the budget implications of the NAPHA policy, and can the Thai government afford them? To answer those questions, we apply the cost structure described above to the baseline scenario (A) and the NAPHA scenario (D1).

Projected Costs (Baseline and NAPHA)

Figure 4.14 displays the financial implications of the NAPHA policy with second-line therapy compared with those of the baseline. We project costs in millions of U.S. dollars and show that they soon attain

Figure 4.14 Projected Net Costs of ART under NAPHA



Source: Authors.

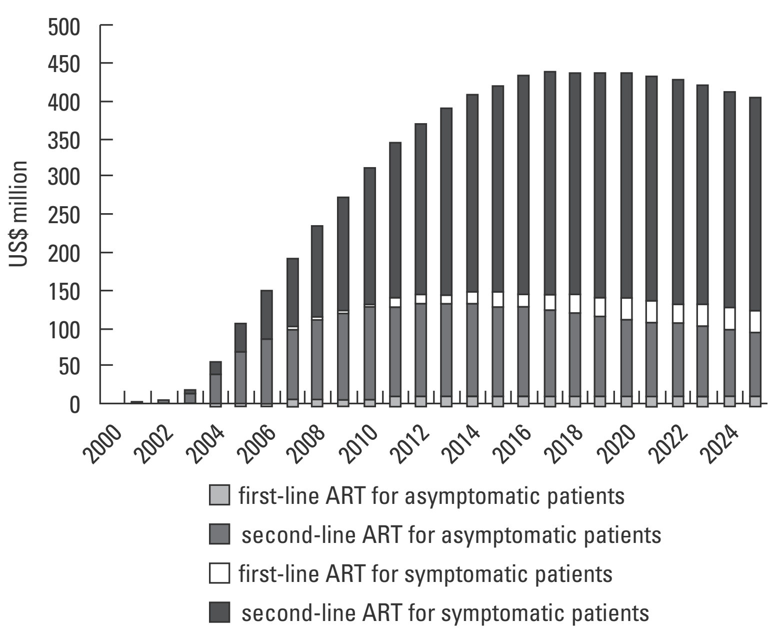
a sum more than 30 times the cost of the baseline scenario. Projected costs of ART are the net costs, which add the costs of ART drugs, other hospital services, VCT, and the net cost of treating opportunistic illnesses. These net cost computations subtract the forgone costs of treatment of those illnesses at the beginning of ART and then add those costs back to the total when treatment fails. Costs will peak in 2019 at US$500 million in 2004 dollars (B 20 billion) per year before leveling off.

Figure 4.15 breaks the total cost of public ART in the NAPHA scenario into four components:

- symptomatic patients in first-line therapy
- symptomatic patients in second-line therapy
- asymptomatic patients in first-line therapy
- asymptomatic patients in second-line therapy.

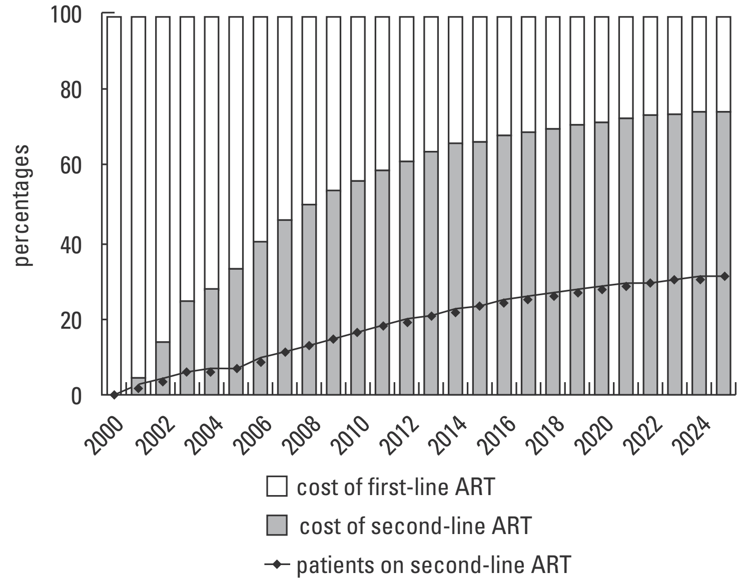
Beginning in 2010, expenditures on second-line therapy generate more than half the cost of treatment. By the end of the projection, second-line therapy for one-fourth of all the patients are responsible for three-fourths of the treatment budget (figure 4.16).

Figure 4.15 Projected Total Cost of Public ART under NAPHA by Line of Treatment



Source: Authors.

Figure 4.16 Percentage of Total Cost of NAPHA Attributable to Second-Line ART



Source: Authors.

As explained above, these projections assume constant future prices. Of course, if the price of second-line treatment falls as much as the price of first-line treatment has fallen, the cost of second-line treatment—and thus the overall costs—will be greatly reduced. However, ongoing World Trade Organization negotiations and bilateral negotiations between the United States and countries such as Brazil, China, South Africa, and Thailand are tending to strengthen patent protection regimes for all drugs. Accordingly, in these simulations, we keep drug prices unchanged throughout the projection period. In chapter 6, we analyze the sensitivity of the cost projections and the cost-effectiveness results to various alternative assumptions, including the possibility of low-cost second-line drugs.

Financial Implications of NAPHA

The calculations above project the annual incremental cost of the NAPHA program to the Thai government.[25] Cost projections show that under the NAPHA policy as modeled here, costs increase sharply, tripling current ART expenditures in a few years. The trend of rapid growth continues until 2020, when total expenditure levels off at close to US$500 million in 2004 dollars (B 40 billion) per year. The present value of the future stream of additional expenditures engendered by

NAPHA totals US$5.6 billion or B 226.4 billion through 2025 in 2004 dollars.

Will expenditures of this magnitude be affordable for Thailand? *Affordable* has no precise economic definition. Affordability is ultimately a political judgment of whether relevant stakeholders will tolerate a given level of spending. However, in almost any country, a sudden and dramatic increase in the amount and proportion of government spending devoted to a single disease requires justification. The cost increases charted in the previous section are certainly sudden and dramatic by any standard.

Tables 4.7 and 4.8 compare those projected future expenditures with the projected levels of the AIDS budget and the total health budget, assuming those two budgets grow at a constant 5.2 percent per year in real terms. In 2004, the net cost of ART under the NAPHA program was estimated at US$77.3 million (B 3,000), 26 percent more than was originally budgeted for AIDS that year and 6.5 percent of the national health budget. From consuming about 1 percent of the health budget and about a third of the AIDS budget in 2001, the public cost of AIDS treatment is projected to rise by 2013 to almost one-fourth of the entire health budget and about seven times the projected (constant growth) AIDS budget. Subsequently costs grow more slowly, so that by 2025 the cost of ART will be about four times the constant growth AIDS budget and about 13.5 percent of the projected health budget.

In addition to general revenue, Thailand has two other sources of revenue to finance the rapidly increasing cost of ART. First, the Global Fund to Fight AIDS, Tuberculosis, and Malaria (GFATM) has already granted the country US$200 million (B 8 billion) for the next few years, of which US$13.5 million (B 540 million) is designated for the MOPH, to cover the cost of the ART program in 2004. Second, several national health insurance schemes—the Universal Coverage Scheme (UCS); the Social Security Scheme for formal sector workers; and the Civil Service Medical Benefit Scheme—are also a source of financing for PHA care, including ART. In 2001, Thailand achieved universal coverage of health care by extending insurance coverage through the UCS. As of 2004, the UCS covers the costs of treatment of opportunistic infections, some lab tests, and hospital services (except the costs of antiretroviral drugs for ART patients, which are

Table 4.7 Projected Costs of the NAPHA Policy for ART, Including Second-Line Therapy, 2001–25

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Baseline													
Net cost of ART	—	10.7	11.4	12.4	12.5	12.4	12.4	12.4	12.4	12.2	12.0	11.6	11.1
First-line ART cost	—	3.3	4.0	4.2	4.3	4.3	4.3	4.4	4.4	4.5	4.6	4.7	4.7
Second-line ART cost	—	0.4	0.6	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.7
Share of second-line cost (%)	—	3.5	5.6	6.1	6.3	6.3	6.3	6.2	6.1	6.1	6.2	6.4	6.6
Share of people on second-line ART (%)	—	1.6	2.2	2.3	2.5	2.5	2.5	2.4	2.4	2.3	2.3	2.2	2.2
NAPHA policy													
Net cost of ART	—	10.9	19.2	33.3	77.3	130.8	176.8	224.0	270.5	314.8	355.4	391.2	421.9
First-line ART cost	—	3.3	10.3	18.9	46.4	78.8	96.6	112.0	125.0	135.5	143.2	148.0	150.5
Second-line ART cost	—	0.5	2.7	8.3	21.6	43.3	71.1	102.4	135.5	169.0	201.8	232.7	260.9
Share of second-line cost (%)	—	4.8	14.3	24.8	27.9	33.1	40.2	45.7	50.1	53.7	56.8	59.5	61.9
Share of people on second-line ART (%)	—	2.3	3.7	5.9	6.3	6.9	9.1	11.1	13.0	14.7	16.4	18.0	19.5

Table 4.7 Continued

	2013	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*	*2023*	*2024*	*2025*
Baseline													
Net cost of ART	10.7	10.3	9.9	9.6	9.4	9.1	8.8	8.5	8.2	7.9	7.5	7.3	7.0
First-line ART cost	4.8	4.9	5.0	5.1	5.1	5.1	5.1	5.1	5.0	4.8	4.7	4.6	4.4
Second-line ART cost	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.6	0.6	0.5
Share of second-line cost (%)	6.9	7.2	7.4	7.7	7.9	8.1	8.2	8.3	8.3	8.3	8.1	8.0	7.8
Share of people on second-line ART (%)	2.2	2.1	2.1	2.1	2.0	2.0	2.0	2.0	1.9	1.9	1.9	1.8	1.8
NAPHA policy													
Net cost of ART	447.3	467.7	483.4	494.8	502.4	506.7	508.1	507.0	503.8	498.9	492.7	485.3	477.0
First-line ART cost	151.0	150.1	148.1	145.3	142.0	138.4	134.5	130.6	126.6	122.6	118.8	115.0	111.4
Second-line ART cost	285.9	307.3	325.0	339.3	350.2	358.2	363.4	366.3	367.1	366.1	363.7	360.0	355.3
Share of second-line cost (%)	63.9	65.7	67.2	68.6	69.7	70.7	71.5	72.2	72.9	73.4	73.8	74.2	74.5
Share of people on second-line ART (%)	20.9	22.2	23.5	24.6	25.6	26.6	27.4	28.2	28.9	29.6	30.2	30.7	31.2

Source: Authors.
Note: All monetary values are in millions of 2004 US dollars. — = not available.

Table 4.8 Projected Costs of the NAPHA Policy for ART, Including Second-Line Therapy, as Percentage of AIDS Budget and Health Budget, 2001–25

	2000	***2001***	***2002***	***2003***	***2004***	***2005***	***2006***	***2007***	***2008***	***2009***	***2010***	***2011***	***2012***
Baseline													
Net cost of ART	—	10.7	11.4	12.4	12.5	12.4	12.4	12.4	12.4	12.2	12.0	11.6	11.1
AIDS budget	—	37.2	36.8	29.7	61.4	67.2	69.5	71.8	49.9	52.5	55.2	58.1	61.1
Health budget	—	1,130.7	1,130.7	1,130.7	1,190.6	1,253.7	1,320.1	1,390.1	1,463.8	1,541.3	1,623.0	1,709.1	1,799.6
Share of AIDS budget (%)	—	28.9	31.0	41.7	20.3	18.5	17.9	17.3	24.8	23.3	21.8	20.0	18.2
Share of health budget (%)	—	0.9	1.0	1.1	1.0	1.0	0.9	0.9	0.8	0.8	0.7	0.7	0.6
NAPHA policy													
Net cost of ART	—	10.9	19.2	33.3	77.3	130.8	176.8	224.0	270.5	314.8	355.4	391.2	421.9
AIDS budget	—	37.2	36.8	29.7	61.4	67.2	69.5	71.8	49.9	52.5	55.2	58.1	61.1
Health budget	—	1,130.7	1,130.7	1,130.7	1,190.6	1,253.7	1,320.1	1,390.1	1,463.8	1,541.3	1,623.0	1,709.1	1,799.6
Share of AIDS budget (%)	—	29.2	52.1	112.1	125.9	194.6	254.5	312.0	542.1	599.6	643.5	673.3	690.2
Share of health budget (%)	—	1.0	1.7	2.9	6.5	10.4	13.4	16.1	18.5	20.4	21.9	22.9	23.4

Table 4.8 Continued

	2013	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*	*2021*	*2022*	*2023*	*2024*	*2025*
Baseline													
Net cost of ART	10.7	10.3	9.9	9.6	9.4	9.1	8.8	8.5	8.2	7.9	7.5	7.3	7.0
AIDS budget	64.3	67.6	71.2	74.9	78.8	82.8	87.2	91.7	96.5	101.5	106.7	112.3	118.1
Health budget	1,895.0	1,995.4	2,101.2	2,212.6	2,329.8	2,453.3	2,583.3	2,720.3	2,864.4	3,016.2	3,176.1	3,344.4	3,521.7
Share of AIDS budget (%)	16.6	15.2	14.0	12.9	11.9	10.9	10.1	9.3	8.5	7.7	7.1	6.5	5.9
Share of health budget (%)	0.6	0.5	0.5	0.4	0.4	0.4	0.3	0.3	0.3	0.3	0.2	0.2	0.2
NAPHA policy													
Net cost of ART	447.3	467.7	483.4	494.8	502.4	506.7	508.1	507.0	503.8	498.9	492.7	485.3	477.0
AIDS budget	64.3	67.6	71.2	74.9	78.8	82.8	87.2	91.7	96.5	101.5	106.7	112.3	118.1
Health budget	1,895.0	1,995.4	2,101.2	2,212.6	2,329.8	2,453.3	2,583.3	2,720.3	2,864.4	3,016.2	3,176.1	3,344.4	3,521.7
Share of AIDS budget (%)	695.6	691.4	679.3	660.9	638.0	611.6	582.9	552.9	522.3	491.7	461.5	432.1	403.7
Share of health budget (%)	23.6	23.4	23.0	22.4	21.6	20.7	19.7	18.6	17.6	16.5	15.5	14.5	13.5

Source: Authors.

Note: All monetary values are in millions of 2004 US dollars. An annual growth rate of 5.2 percent is assumed for the AIDS budget, health budget, and public health spending from 2004 to 2025, on the basis of 2004 figures. AIDS budgets during 2004 and 2008 include first and second round allocations from the Global Fund for AIDS, Tuberculosis, and Malaria.

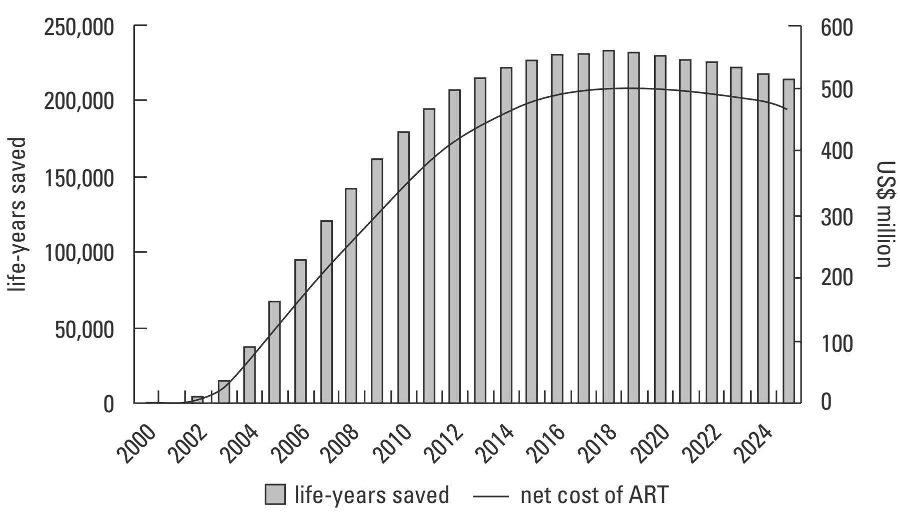
covered by NAPHA), with a nominal fee of less than US$0.80 per visit (B 30). The Social Security Scheme covers treatment of opportunistic infections and, since August 2004, also the cost of ART (with caps) for its affiliated members (but not for their families). The Civil Service Medical Benefit Scheme covers all care for PHAs including ART (see chapter 2).

Although the projection shows a significant increase in ART expenditure over time, it remains less than 24 percent of the (constant growth) national health budget. Because a reduction of more than 20 percent in other health expenditures is unlikely, Thailand can finance these additional expenditures from such outside sources as the GFATM and health insurance schemes, or it can increase the size of the overall health budget as a share of total government spending.

Cost-Effectiveness of NAPHA

Figure 4.17 presents the model's projections of the annual benefits and the annual costs of the NAPHA policy, compared with the baseline scenario of limited public sector access to ART. The annual benefits measured in life-years saved increase rapidly, peaking in 2017. These benefits begin to level off as the stock of AIDS patients gradually declines. Because of the slow transition of patients to the more expensive second-line therapy, costs continue rising, peaking at about US$500 million (B 20 billion) in 2020 before leveling off.

Figure 4.17 Benefit (Life-Years Saved) and Costs of NAPHA Relative to Baseline

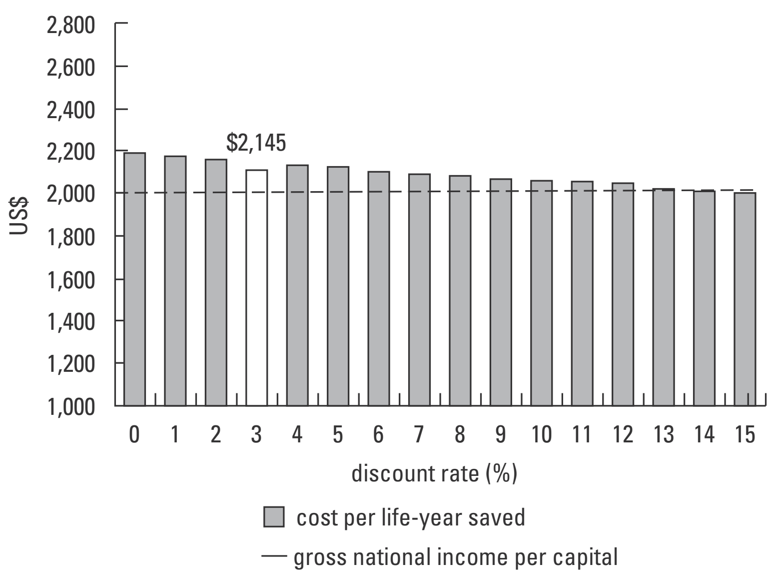


Source: Authors.

When the benefits of any intervention occur relatively soon after the program is launched and costs occur later, discounting reduces the importance of the future costs in relation to the near-term benefits and thus improves the cost-effectiveness of the program. Figure 4.18 presents the net present value of the cost per life-year saved for the NAPHA policy, with all costs and benefits discounted back to 2002. The choice of discount rate has a small but significant effect on the result. At a zero discount rate, the NAPHA program costs close to US$2,200 (B 88,000) per life-year saved, whereas at a 15 percent discount rate, the cost of the same life-year saved would be US$2,000 (B 8,000). At the conventional discount rate of 3 percent, the cost per life-year saved for the NAPHA policy is US$2,145 (B 85,800), slightly greater than Thailand's 2002 gross national income per capita of US$2,000 (B 80,000).

Is US$2,145 (B 85,800) a large or a small price for the Thai government to pay to save years of life through a special treatment subsidy program? Different observers have differing views. An argument for horizontal equity would ask whether the government is willing to pay as much to buy life-years through subsidized treatment of other adult illnesses, such as cancer, heart disease, or end-stage renal disease. Advocates of vertical equity would point out that this sum is small enough to make treatment a good personal investment for people in

Figure 4.18 Cost Effectiveness of NAPHA at Different Discount Rates



Source: Authors.

the upper quintile of Thailand's income distribution, where incomes per capita exceed US$10,000 (B 400,000). Therefore, in the interest of vertical equity, the government should ensure that people in the bottom four-fifths of the income distribution have access to care that the those in the top one-fifth will purchase for themselves. Before entering further into this discussion (in chapter 6), we turn to the question of how modifications of NAPHA will affect its benefits and costs, possibly improving its attractiveness as a health policy.

Notes

1. Christianson (1976) is one of the earliest examples. See the other sources cited for applications in a development context. None of these authors considers the demand for diagnostic information as a precursor to the demand for care, as we do in the discussion of figure 4.2.

2. In microeconomic theory, this assumption is referred to as homotheticity and is frequently adopted to improve the tractability of economic models of demand.

3. McFadden (1974) first used the conditional multinomial logit specification to model the commuters' choice between bus, car, and train modes of transportation in the San Francisco Bay area.

4. Note that the numerator of the multinomial logit demand function is an exponential function of price and distance. Exponential functions are always positive. Further note that the denominator of the demand function is the sum of the numerators for the four modes. Hence, by the choice of this functional form, each mode demand is guaranteed to be a positive fraction between zero and one, and the sum of the four demand functions is guaranteed to equal one.

5. To see this relationship, one may note that according to the functional form, the logarithm of the ratio of two proportions—say public and augmented public—can be written as a linear function of the logarithms of the price and distance ratios between the same two options: $LN(q\_pub\_art/q\_apub\_art) = (\alpha\_pub - \alpha\_apub) + \beta * LN(P\_pub\_art/P\_apub\_art) - \gamma * LN(D\_pub\_art/D\_apub\_art)$.

6. Economists refer to a demand sensitivity such as this, where the response is smaller than the stimulus, as inelastic. In microeconomic studies of the demand for acute medical care, patients have typically been found to be inelastic to price, with elasticities of –0.2 or even closer to zero. We have chosen a value less close to zero because we are modeling the behavior of a population that is on average quite poor and therefore can be expected to be more sensitive to price.

7. When a lower price for one good increases the demand not only for it but also for another good, the two goods are said to be complements in consumption.

8. We consider that every HIV-negative adult has a probability of seeking VCT in any given year.

9. This is the meaning of the term semiempirical, which the published papers use to describe the model.

10. Although sometimes referred to as migration, the behavior captured by these assumptions primarily concerns sexual behavior rather than geographic migration. In the case of sex workers, the transition out of sex work may often be accompanied by a physical change of location, such as a return to a rural hometown. The AEM does not attempt to model geographic location or movements.

11. An example of the difference equations that are the building blocks of the model is the expression for new infections among male clients (Thai Working Group on HIV/AIDS Projection 2001).

12. The figure is not large enough to display the full structure for both early and late treatment. The reader is asked to imagine the same structure for late treatment as is displayed for early treatment.

13. Because we generally assume that Thai policy on the availability of ART will accommodate demand, this constraint on treatment slots is rarely binding in the model runs presented here.

14. The study of the costs and effects of AIDS treatment in India by Over and others (2004) presents the discounted averted years of

orphanhood estimated to result from the simulated alternative policies. However, it eschews the attempt to aggregate these benefits with the benefits of healthy life-years saved.

15. The rate of discount is much less certain than the fact that discounting is necessary. One golden rule is that costs and effects should be discounted at the same discount rate. However, the best choice for that rate is not clear. This report follows the lead of the 1993 World Development Report on health in adopting a discount rate of 3 percent.

16. We assume that the private sector will serve only symptomatic patients. Although not strictly correct, this assumption may approximate reality to the extent that demand by the asymptomatic is more price elastic.

17. The augmented public and private sector demands are not shown in the figure.

18. Here we discuss the baseline and NAPHA scenarios, returning in the next section to the VCT scenario.

19. AEM baseline data show approximately 46 million low-risk adults in 2001 and about 2.6 million high-risk adults.

20. The far right demand curve in panel (b) illustrates the effect of the expanded VCT scenario, discussed in chapter 5.

21. This section draws on Masaki (2004).

22. We assume that HIV-infected people under treatment are 75 percent less infectious on each sexual contact than they would have been without ART. Chapter 6 performs sensitivity analysis with respect to this assumption.

23. In a real population, this measure would be increased by both the number of people who are living longer because of effective ART and also by the fertility of those people. However, the AEM does not model

fertility. Instead it grows the population by adding a new 15-year-old cohort of people every year, whose size is independent of the size of the current population. Therefore, the difference between two scenarios in the size of the population in any year is due entirely to differences in ART.

24. We do not adjust life-years for the degree of disability as was done in computing the disability adjusted life-years burden of disease in the 1993 World Development Report (World Bank 1993).

25. This section is based on Masaki (2004).