METHODS Appendix

This appendix provides a detailed description of the methods used in the article Cost-Effectiveness of Antiretroviral Regimens in the World Health Organization’s Treatment Guidelines: A South African Analysis (Bendavid, et al.)

Overview

We developed a simulation model of HIV disease that followed the natural history of HIV-infected individuals from the time of presentation to care until death. Taking a societal perspective, the model follows the costs and benefits of five treatment strategies in sub-Saharan Africa (four recommended by the World Health Organization – WHO – and the ART combination in most common use in sub-Saharan Africa). The model estimates discounted and undiscounted quality-adjusted life expectancy from the time of presentation to care until the end of life in 2009 US dollars. The comparative value of alternative treatment strategies is expressed in terms of incremental cost-effectiveness ratios of each strategy compared with the next less effective strategy.

Model structure

The model follows patients in one month intervals from the time at which they first make contact with the medical system for their HIV infection until death. Figure 1 shows a schematic flow diagram of patient care tracked in the model. At presentation and at each clinic visit, patients who are not on ART are evaluated whether they meet criteria to start ART. ART is initiated when CD4 cell count drops below 350 cells/μl, as suggested in the recent WHO guidelines. Patients who are on first-line ART also arrive to the clinic in regular intervals, and are evaluated for continued efficacy and possible toxicities of first-line ART. Patients are switched from first-line ART to second-line ART for two reasons: suspected treatment failure (by immunologic or clinical criteria) and drug toxicity severe enough to necessitate a change in regimen. Patients on second-line ART are also evaluated at regular intervals for signs of medication toxicity and treatment of opportunistic diseases (ODs). Second-line treatment is stopped in patients who have severe medication toxicity or are otherwise unable to tolerate the medications. However, patients who are on second-line ART and have evidence of virologic failure are maintained on ART due to the independent survival advantages of non-suppressive regimens compared with ART cessation.[1, 2]

Patients may present at any clinical stage of disease and with any laboratory parameters. That is, on entry, a patient may or may not be ill with an opportunistic disease and may have any CD4 count or viral load. The distribution of CD4 counts at entry was taken from published cohort studies in the Cape Town region, and their risk of presenting with an OD was dependent on their CD4 count at presentation.[3, 4] Individuals with CD4 counts under 350 cells/μl are placed on treatment at the beginning of the simulation. Those with higher CD4 counts are observed until their CD4 count dropped below the threshold; in practice, very few individuals present with CD4 counts higher than 350 cells/μl. We assume that no patients have transmitted drug resistance (TDR). Although TDR does impact the treatment efficacy of NNRTI-based regimens, the rates of TDR reported from areas of Africa using the consensus definition of TDR have generally been low and thus are unlikely to have a major impact on treatment efficacies beyond the observed rates of virologic failure.[5, 6] Estimated rates of virologic suppression for each regimen are displayed in Table 1.

The model evaluates all patients in one month intervals. While the model tracks all patient parameters including CD4 count, viral load, ART regimen, medication toxicities, and development of ODs – most parameters are only available to providers during regular clinic visits. That is, while the model tracks an individual’s health status monthly, that patient’s data is only available for treatment decisions if it is measured and if that patient presents to clinic that month.

If a patient experiences a severe OD, we assume they present for acute medical care rather than to a routine clinic visit that month. The risk of developing a severe OD was dependent on the current CD4 count. We calculated the risk of a severe OD based on the risk of developing most WHO Stage 4 diseases (CMV infection, cryptococcal meningitis, Toxoplasmosis, Pneumocystis pneumonia, extrapulmonary TB, wasting, esophageal candidiasis, and chronic diarrhea) plus the risk of pulmonary TB based on experience in Cape Town.[3, 4, 7] We used data from patient cohorts that received cotrimoxazole prophylaxis.

A patient’s risk of death from an OD is proportional to the CD4 count at the time of illness. The costs incurred reflect care for the OD. If a patient survives the acute illness, he/she returns to routine care. Patients were followed until death from HIV or other causes (background age-specific mortality rate).[8–11] Thus, we follow the lifetime costs and benefits of HIV care delivery for a group of simulated patients using clinical and utilization data of cohorts.
Disease progression

We follow the disease progression of patients from the time of presentation based on the following parameters: age, CD4 count, viral load, ART regimen, ART duration, history of OD, virologic failure, and medication toxicity. We monitored all parameters for all patients monthly, but the information was only available to providers every 6 months or sooner for acute clinical events (such as onset of an OD or medication toxicity).

Upon entry to care, each patient is assigned an initial CD4 count, viral load, and age from a distribution that is calibrated to Cape Town study cohorts.[4, 12–14] Each patient’s risk of clinical events is determined by his/her CD4 count. The CD4 count was modeled as a continuous variable that varied based on the viral load, ART, and occurrence of treatment failure. In patients whose viral load was not suppressed, the rate of CD4 decline was determined by their current CD4 count and viral load.[15, 16] Given the uncertainty about the exact relationship between viral load and CD4 change, we allow two non-linear determinants of CD4 decline: random variability that loosens the correlation between viral load and CD4 count decline, and a slower rate of CD4 decline, both guided by published data.[15–17]

Once a patient is started on a successful first-line ART regimen, his/her CD4 rises to a peak that depends primarily on the CD4 count at the time of treatment initiation. While some data support an age-related effect of CD4 rise, the strongest reproducible predictor of CD4 rise on effective ART is the CD4 count at the time of treatment initiation.[18–21] Published data on CD4 rise were extracted using the graph digitizing program DigitizeIt v.1.5.8 (Braunschweig, Germany), and monthly CD4 increments were determined based on time elapsed from treatment initiation.

The principal activity of ART is suppression of viral replication, and we use viral suppression to undetectable levels as the principal marker that allows CD4 to rise after treatment initiation. While on successful treatment, viral load is undetectable at a threshold of less than 400 copies/ml.

Treatment failure is modeled as failure to suppress virologic replication and a return of the viral load to detectable levels. Patients with virologic failure who are continued on ART have a lower viral “set point,” and their rate of CD4 decline is consequently slower.[2] Clinically, virologic failure is inferred through CD4 monitoring, and we use immunologic criteria outlined in the WHO guidelines and recent clinical trials—a drop to a CD4 count of less than 100 cells/μl—to estimate timing of virologic failure and the need to switch to second-line therapy.[7, 22, 23]

Treatment strategies

Regimens

We compared the effectiveness and costs of five alternative ART first-line regimens (four recommended by the WHO and the ART combination in most common use in sub-Saharan Africa):

1. Tenofovir + lamivudine + efavirenz
2. Tenofovir + lamivudine + nevirapine
3. Zidovudine + lamivudine + efavirenz
4. Zidovudine + lamivudine + nevirapine
5. Stavudine + lamivudine + nevirapine

All the regimens have a similar purpose—to suppress viral replication and enable immunologic recovery—but they differ substantively on two primary domains: success rates of achieving virologic suppression and their respective toxicity profile. Table 1 shows the estimates of each regimen’s rates of virologic suppression and toxicity profile used in the model.

Virologic suppression

We estimate rates of virologic suppression from comparative trials. Using data from clinical trials performed mostly in developed country settings raises questions about generalizability to an African setting. We use this data for two primary reasons. First, it is the best available comparative data for these regimens. Literature with estimates of virologic suppression from uncontrolled cohorts suggests that data are scant and unreliable, as it fluctuates widely within regimens based on the patient population, virologic assay, and case definitions for virologic failure. Second, recent literature suggests that suppression rates are similar between subtype B and non-subtype-B, despite the genotypic differences, supporting observations that ART effectiveness is similar between developed and developing settings.[24]

We rely heavily on two long-standing clinical trials in establishing rates of virologic suppression. One compares regimens containing tenofovir to regimens containing stavudine, while the other compares regimens containing tenofovir to regimens containing zidovudine.[25, 26] Those studies show that tenofovir and stavudine are similarly efficacious, while tenofovir is more efficacious than zidovudine. Since both studies were performed using a similar protocol with the same group of investigators, we assume by transitivity that stavudine is more efficacious than zidovudine. Both studies use efavirenz as the NNRTI of choice. We estimated the rates of virologic failure with nevirapine were about 1.5 times...
to maintain viral suppression in combination with a variety of NRTIs.[27–31]

**Toxicities and regimen changes**

We include the effect of seven dominant toxicities associated with ART: lipoatrophy, renal failure, anemia, hepatotoxicity, myocardiopathies peripheral neuropathy, and lactic acidosis. While other toxicities are known to be associated with ART, we chose to examine those toxicities that are most common and significant in terms of their effect on quality of life. Where possible, we estimate the types of toxicities and incidence rate for each regimen from long-term follow-up studies of clinical trials.[26, 32, 33] We use these sources because of the strict case definitions and careful monitoring. Where clinical trial data was not available, we use African observational data.[30, 34–37] In particular, we rely on cohorts that identified toxicities that led to regimen change as the clinically relevant endpoint. For a few toxicities we rely on observational data from non-African cohorts.[38, 39] Table 1 shows the toxicities associated with each regimen and 1-year frequency of each toxicity. The cumulative risk of all the toxicities except for lipoatrophy plateaus after a year, and individuals on regimens associated with each toxicity who remain on the regimen for at least a year without experiencing the toxicity are no longer at risk. The most common toxicity, associated with all the regimens, is lipoatrophy. It occurs most frequently with stavudine-based regimens and least frequently with tenofovir-based regimens. We estimate a declining rise in the risk of lipoatrophy up to three years, when the cumulative risk plateaus.

Therapeutic decisions following toxicities aim to minimize the risk of future toxicity burden. Practically, most toxicities that are associated with zidovudine or stavudine prompt a switch to a tenofovir-containing regimen. Most toxicities associated with nevirapine prompt a switch to an efavirenz-containing regimen. The main exception is renal failure with tenofovir, which prompts a switch to a zidovudine-containing regimen. These therapeutic decisions are shown in Figure 1 of the main manuscript. These regimen substitutions with toxicities are based primarily on the WHO formulary.

Consistent with WHO guidelines and with standard practice in many parts of sub-Saharan Africa, we modeled a second-line ART for those who experience toxicities on multiple first-line regimens or who are thought to fail first-line ART. Second-line ART included a combination of NRTIs that depended on the initial regimen and a boosted protease inhibitor.[7]

**Benefits and costs**

Benefits are measured in life years (LYs) and quality-adjusted life years (QALYs) from the time of presentation. We compared both discounted and undiscounted life years and QALYs among the various treatment strategies. Where possible, we used quality of life estimates from the same clinical trials that reported the toxicity incidence. We had this information for neuropathy. We use a study on switching from stavudine to tenofovir in South Africa for most other quality of life estimates. That study, in turn, uses a general quality of life catalog to estimate many of the associated weights. Because of the uncertainty associated with the QALY weights, and their importance in shaping the results, we varied the quality of life weight estimates for each toxicity widely (Table 2 in main manuscript).

We consider direct costs of care from a societal perspective in this study. We included the costs of inpatient care, outpatient care, provision of ART, laboratory monitoring, and treating toxicities. Inpatient and outpatient clinic costs are taken from a detailed costing study of HIV care in South Africa.[40, 41] The differences in cost of care between South Africa and other sub-Saharan countries poses a legitimate concern to the study’s generalizability. Consequently, we varied the costs widely based on measured variations in cost of medical care provided in the WHO-CHOICE database.[42]

We obtained the cost of each regimen from the WHO Global Price Reporting Mechanism database.[43] That database provides price data of antiretroviral drugs obtained directly from national AIDS programs and other major purchasers, and publishes detailed transactional information, including quantities purchased, dosages, and the amount paid. Some antiretroviral drugs have fixed-dose combinations, which generally reduce the price of the regimen. For example, stavudine, lamivudine, and nevirapine come in a fixed-dose combination, as do zidovudine and lamivudine. Regimens with fixed-dose combinations are cheaper than regimens with the individual drugs procured independently. We obtained the lowest price for each regimen reported in the Global Price Reporting Mechanism for South Africa in 2008. Notably, the prices of all ART regimens converged in sub-Saharan Africa by 2008, and the prices paid for ART were nearly identical between South Africa and other African countries.[44]

**Sensitivity analysis**

Our sensitivity analysis includes several one-way and multi-way analyses, as well as a probabilistic sensitivity analysis. The uncertainty bounds for each parameter are shown in Table 1 of the main manuscript. We pay particular attention to those data elements for which we have less certainty and which change significantly between alternative strategies. These include the rates of failure, toxicity rates, quality of life weights with toxicities, and cost of the antiretroviral agents. We perform a probabilistic sensitivity analysis where we vary all the variables simultaneously and repeat the analysis.
1,000 times. Each variable is drawn from a probability distribution and the entire analysis is re-run. Probabilities for events or health states were sampled using beta distributions with alpha and beta parameters determined by the point estimate (mean) and variance; and costs were sampled using gamma distributions with the mode at the point estimate. Beta distributions are defined by two shape parameters, \( \alpha \) and \( \beta \), that were estimated for each variable to approximate the mean and variance as follows:

\[
\alpha = (\mu^2 - \mu^3 - \mu \sigma^2)/\sigma^2
\]

\[
\beta = (\mu - 2 \mu^2 + \mu^3 - \sigma^2 + \mu \sigma^2)/\sigma^2
\]

Gamma distributions are defined by a shape and scale parameters, \( k \) and \( \theta \):

\[
k = \mu^2 / \sigma
\]

\[
\theta = \sigma / \mu
\]

This allows us to estimate the confidence in our results if the true value of each variable is anywhere within the uncertainty bounds shown. For example, we estimate the likelihood that a strategy which appears dominant in the base case – one which is more effective and less costly than another strategy – may not be dominant.

### References


2. Ledergerber, B., et al., Predictors of trend in CD4-positive T-cell count and mortality among


42. CHOosing Interventions that are Cost Effective (WHO-CHOICE). September 27, 2008; Available from: http://www.who.int/choice/en/.


