The Effect of Monitoring Viral Load and Tracing Patients Lost to Follow-up on the Course of the HIV Epidemic in Malawi: A Mathematical Model

Janne Estill,1,2, Cliff C. Kerr,3, Nello Blaser,4, Luisa Salazar-Vizcaya,3,4, Lyson Tenthani,3,4, David P. Wilson,4, and Olivia Keiser1,3

1Institute of Global Health, University of Geneva, Geneva, Switzerland; 2Institute of Mathematical Statistics and Actuarial Science (IMSA) and 3Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland; 4Burnet Institute, Melbourne, Australia; 5School of Physics, University of Sydney, Sydney, Australia; 6Department of Mathematics, University of Bergen, Bergen, Norway; 7Department of Infectious Diseases, University Hospital Bern, Bern, Switzerland; 8Supporting Operational AIDS Research (Project SOAR), Population Council, Blantyre, Malawi

Background. Antiretroviral therapy (ART) reduces HIV transmission, but treated patients may again become infectious. We used a mathematical model to determine whether ART as prevention is more effective if viral load (VL) is routinely monitored and patients lost to follow-up (LTFU) traced.

Methods. We simulated ART cohorts to parameterize a deterministic transmission model calibrated to Malawi. We investigated the following strategies for improving treatment and retention: monitoring VL every 12 or 24 months, tracing patients LTFU, or a generic strategy leading to uninterrupted treatment. We tested 3 scenarios, where ART scale-up continues at current (Universal ART), reduced (Failed scale-up), or accelerated speed (Test&Treat).

Results. In the Universal ART scenario, between 2017 and 2020 (2050), monitoring VL every 24 months prevented 0.5% (0.9%), monitoring every 12 months prevented 0.8% (1.4%), tracing prevented 0.3% (0.5%), and uninterrupted treatment prevented 5.5% (9.9%) of HIV infections. Failed scale-up resulted in 25% more infections than the Universal ART scenarios, whereas Test&Treat resulted in 7%–8% less.

Conclusions. Test&Treat reduces transmission of HIV, despite individual cases of treatment failure and ART interruption. Whereas viral load monitoring and tracing have only a minor impact on transmission, interventions that aim to minimize treatment interruptions can further increase the preventive effect of ART.

Keywords. antiretroviral therapy; HIV; loss to follow-up; mathematical model; monitoring; transmission.
from untreated patients, behavioral preferences, and contact patterns. In this study, we took the next step and developed a transmission model to assess the potential effect of VL monitoring and tracing on the HIV epidemic.

**METHODS**

**Setting**

We modeled the HIV epidemic of Malawi. In 2016, the estimated adult HIV prevalence in Malawi was 9.2%, and about 680,000 patients (65% of all PLHIV) were on ART [14]. Until recently, Malawi relied on clinical monitoring and occasional CD4 counts to monitor treatment response, but since 2011, the Ministry of Health has recommended monitoring VL at regular 24-month intervals [15]. In 2016, 19% of all patients on ART in Malawi had had at least 1 VL test. Several sites trace patients who miss appointments [16].

**Mathematical Model**

The model consists of an individual-based simulation of disease progression and a deterministic transmission model.

**Disease Progression Simulation**

We used the R package *gems* to develop an individual-based simulation model for disease progression in patients who have started ART ("treatment model") [17, 18]. We divided the patient’s time on ART into 32 states (Figure 1A; Supplementary Table 1) that accounted for virological and immunological treatment response (successful or failing), ART regimen (first- or second-line), and retention (on or off ART). We did not consider further treatment options beyond second-line. We separated HIV-related and HIV-unrelated mortality. We limited the number of off-ART episodes to 1 to simplify the structure. Patients were simulated for 10 years after initiating ART. Virological and immunological treatment response was based on previous analyses of routine data from sites in South Africa with 6-monthly VL and CD4 monitoring (Supplementary Table 2) [12, 19, 20]. The risk of virological failure corresponds to a cumulative risk of 5.7% 1 year and 12.9% 5 years after ART initiation. One year after ART initiation, the sensitivity of immunological criteria was 7% and specificity 12%; 5 years after ART initiation, the sensitivity and specificity were 26% and 45%, respectively. A resistance penalty factor was applied to increase the risk of failure depending on how long the patient had spent on a failing regimen or off ART. Parameters related to treatment interruptions and return to care with or without tracing were derived from data from Lighthouse and Martin Preuss Centre clinics in Lilongwe, Malawi [13]. Patients switched to second-line ART after either virological or immunological treatment failure, depending on the monitoring strategy.

![Figure 1. Schematic representation of the mathematical model. A, Flow of patients in the treatment model. White boxes represent stages with suppressed viral load, and gray boxes represent stages with continuously elevated viral load. "Discordant" immunological failure refers to a decline in CD4 cell count fulfilling the failure criteria under suppressed viral load; this condition will not reverse upon switch to second-line therapy. The flow described on the upper half is applicable to patients on ART, including those who returned after ART interruption. While progressing along the stages of treatment response (upper graph), the patients may also interrupt and restart treatment or die (lower graph). B, Transmission model. The upper graph shows the course of the HIV infection, and the lower graph the flow through age, sex, and risk group. Black arrows show flows between compartments, and gray lines show sexual contact patterns. Abbreviation: ART, antiretroviral therapy.](https://academic.oup.com/ofid/article-abstract/5/5/ofy092/4987343)
The output of the simulations is a matrix of the entry times to all states for each patient. For each simulated patient, we sampled the VL values for different stages (virologically successful, failing, or off ART). We then calculated the mean log10 viral load over time in each strategy to inform the transmission model [12]. We ran the model for the following nine strategies:

i. no VL monitoring, no tracing;
ii. no VL monitoring, active tracing;
iii. no VL monitoring, uninterrupted treatment;
iv. VL monitoring at 24-month intervals, no tracing;
v. VL monitoring at 24-month intervals, active tracing;
vi. VL monitoring at 24-month intervals, uninterrupted treatment;
vii. VL monitoring at 12-month intervals, no tracing;
viii. VL monitoring at 12-month intervals, active tracing;
ix. VL monitoring at 12-month intervals, uninterrupted treatment.

“No VL monitoring” means that treatment failure is again determined by clinical symptoms and CD4 counts only. The WHO recommends VL monitoring at 12-month intervals, but VL monitoring at 24-month intervals is current practice in Malawi. Treatment failure is confirmed 3 months after detection, and the patient is switched to second-line therapy after a random delay. In strategies with no tracing, patients LTFU may only return spontaneously. With active tracing, all patients LTFU are traced 3 weeks after missing their appointment and return to care at a higher rate than without tracing [16]. Strategies with uninterrupted treatment represent the ideal but unrealistic scenario that treatment is never interrupted. This scenario sets the theoretical limit for improving retention. In all scenarios, we assumed that patients retained in care adhered to treatment, as in the data that were used for parameterization [12, 13, 19, 20].

Transmission Model

We developed a deterministic transmission model (“transmission model”) to represent the HIV epidemic in Malawi between 1975 and 2050. The model consists of 40 compartments (Figure 1B) representing HIV status (susceptible; primary, asymptomatic chronic, or symptomatic chronic infection; ART; AIDS), age (children <15 years, adults 15–49 years, adults ≥50 years), sex (not distinguished for children), and risk behavior (high or low, except for children, older adults, and AIDS patients). “Acute infection” represents early HIV infection, when the risk of onward transmission is highest [21]. During “asymptomatic chronic infection,” CD4 cell counts are expected to be >350 cells/μL. In “symptomatic chronic infection,” CD4 cell count is below 350 cells/μL and the patient is in WHO clinical stage ≥2. Although there is no strict dependency between symptoms and clinical stage, these definitions roughly correlate [22]. “ART” represents patients who ever started ART, including also patients who interrupted ART. “AIDS” represents the last year of untreated HIV, where the patient has a CD4 cell count below 200 cells/μL or severe AIDS-defining opportunistic infections.

The model is solved numerically. We calculated infectiousness from the estimated per-act transmission probability and assumed frequency of unprotected sex acts (Table 1). Infectiousness was multiplied by 20 during the acute stage, and by a factor that depended on the monitoring and retention strategy during ART. This factor was estimated for each scenario from the treatment model. We also allowed infectiousness to decrease over time. The progression of patients across the stages of HIV was estimated from the literature.

We calculated the ART initiation rate at based on the diagnosis rate, availability, and eligibility criteria for ART and the expected progression of the patient’s CD4 cell count (Table 1). Antiretroviral therapy was generally unavailable until 2003, provided to symptomatic patients only between 2003 and 2015, and increasingly to asymptomatic patients after 2015. The rate of ART initiation among symptomatic patients increased in 2011 to match the change in CD4-based eligibility criteria (from <250 to <350 cells/μL). A separate ART initiation rate was applied to women from 2011 on to take into account the “Option B+” strategy of treating all pregnant and breastfeeding women [23].

We ran the simulation for 1975–2016 with the best available parameter estimates to calibrate the model and compared the model’s outputs with observed data, including total population size and number of patients on ART each year from 2010. We also compared our results to the estimates of the UNAIDS EPP/Spectrum model on prevalence and number of annual new infections. We then calibrated the following parameters: per-act transmission probability, birth rate, and year and magnitude of decrease in infectiousness.

We considered 3 possible scenarios for treatment access from 2017 (see Table 2 for input parameters):

1) Failed scale-up: Recommendations to treat all PLHIV are not successfully implemented, and treatment remains restricted to the sickest patients. Women can start ART earlier because “Option B+” has already been implemented.

2) Universal ART: The policy introduced in 2015 continues, and ART can be initiated during the asymptomatic stage, although at a lower rate than in symptomatic patients because of barriers to testing.

3) Test&Treat: Intensive screening is added to the Universal ART scenario, and ART initiation rate increases.

We ran each of the 3 access scenarios for all 9 monitoring and retention strategies, (i)–(ix). We reported the number of expected new infections and AIDS-related deaths in 2020, 2030, and 2050 and calculated the relative reduction in new infections.
compared with strategy (i) of the corresponding access scenario for 2 time windows: 2017–2020 and 2017–2050.

**Sensitivity Analyses**

We conducted 2 sensitivity analyses to assess the impact of uncertainty around the parameters (Supplementary Table 3). In the first analysis, we assumed that the risk of virological failure would considerably increase over time. Second, we conducted an analysis where all-cause mortality was reduced from 2017 onwards, and treated patients could no longer proceed to AIDS.

**RESULTS**

The mean (log10 scale) viral load of the patients simulated in the treatment model for 10 years ranged from 63 to 104 copies/mL across the scenarios (Supplementary Table 4). The corresponding per-act transmission risk from patients who started ART was 14–17 times lower than the risk from chronically infected, untreated patients. The per-act transmission probability determined via calibration of the transmission model was 20% higher than the literature-based prior (Supplementary Table 5).
The transmission model’s pre-2017 results were in line with observed data and the UNAIDS EPP/Spectrum predictions. Prevalence among adults aged 15–49 years followed the upper limit of the UNAIDS estimates (Supplementary Figure 1) [14]. The total number of PLHIV in 2016 was about 10% lower than predicted by UNAIDS. The largest discrepancy was in annual new infections, which our model predicted to be about 20% higher until 2013. In the last few years, the new infections declined rapidly, going below the UNAIDS lower limit in 2016. According to our model, 695,000 patients were on ART in 2016, in line with observed estimates (680,000). In 2016, we predicted 23,100 AIDS-related deaths, whereas UNAIDS predicted 24,000.

In all scenarios, prevalence continued to decline. Prevalence among adults aged 15–49 years was 7.8%–7.9% in 2020, 4.4%–4.8% in 2030, and 1.4%–1.7% in 2050, depending on the scenario (Supplementary Figure 2). The number of annual new infections also decreased rapidly, ranging 8800–13,400 in 2020, 6400–9900 in 2030, and 3500–6100 in 2050 across the scenarios (Figure 2, Table 3). The number of AIDS deaths was 12,000 in 2020, 7400–7900 in 2030, and 4600–5400 in 2050.

Of the factors that differed between scenarios, results were most sensitive to the overall treatment and testing scenario. In the Universal ART scenario, our model predicted 50,100–53,400 new infections between 2017 and 2020, or 231,600–260,900 between 2017 and 2050. In the Failed scale-up scenario, the ranges were 56,800–60,200 until 2020 (13% higher than with Universal ART) or 288,900–326,800 until 2050 (25% higher than with Universal ART). With Test&Treat, the ranges dropped to 46,200–49,400 until 2020 or 214,900–241,500 until 2050, about 7%–8% lower than with Universal ART.

The differences between monitoring and retention strategies were smaller (Figure 3, Table 3). Assuming the Universal ART scenario, current retention, no tracing, and no VL monitoring, 53,400 patients were infected between 2017 and 2020, or 260,900 between 2017 and 2050. Monitoring VL at 24-month intervals lowered these numbers by 0.5% to 53,200 until 2020, or by 0.9% to 258,700 until 2050; and, at 12-month intervals, by 0.8% to 53,000 until 2020, or by 1.4% to 257,200 until 2050. The relative differences remained the same across all scenarios of treatment scale-up and retention/tracing.

In the Universal ART scenario without viral load monitoring, actively tracing patients LTFU reduced the average number of new infections in 2017–2020 by 0.3% from 53,400 to 53,300, or in 2017–2050 by 0.5% from 260,900 to 259,500 (Figure 3, Table 3). When treatment interruptions were eliminated, the number of new infections decreased in 2017–2020 by 5.5% to 50,500, or in 2017–2050 by 9.9% to 235,100. The relative benefit of tracing and improved retention was similar across scenarios of treatment scale-up and monitoring.

The total number of PLHIV also decreased over time. In 2020, we predicted 839,200–849,500 PLHIV, and in 2050, 465,300–538,000 PLHIV. Because of the decreasing number of PLHIV, the relative differences remained the same across all scenarios of treatment scale-up and retention/tracing.

The transmission model’s pre-2017 results were in line with observed data and the UNAIDS EPP/Spectrum predictions. Prevalence among adults aged 15–49 years followed the upper limit of the UNAIDS estimates (Supplementary Figure 1) [14]. The total number of PLHIV in 2016 was about 10% lower than predicted by UNAIDS. The largest discrepancy was in annual new infections, which our model predicted to be about 20% higher until 2013. In the last few years, the new infections declined rapidly, going below the UNAIDS lower limit in 2016. According to our model, 695,000 patients were on ART in 2016, in line with observed estimates (680,000). In 2016, we predicted 23,100 AIDS-related deaths, whereas UNAIDS predicted 24,000.

In all scenarios, prevalence continued to decline. Prevalence among adults aged 15–49 years was 7.8%–7.9% in 2020, 4.4%–4.8% in 2030, and 1.4%–1.7% in 2050, depending on the scenario (Supplementary Figure 2). The number of annual new infections also decreased rapidly, ranging 8800–13,400 in 2020, 6400–9900 in 2030, and 3500–6100 in 2050 across the scenarios (Figure 2, Table 3). The number of AIDS deaths was 12,000 in 2020, 7400–7900 in 2030, and 4600–5400 in 2050.

Of the factors that differed between scenarios, results were most sensitive to the overall treatment and testing scenario. In the Universal ART scenario, our model predicted 50,100–53,400 new infections between 2017 and 2020, or 231,600–260,900 between 2017 and 2050. In the Failed scale-up scenario, the ranges were 56,800–60,200 until 2020 (13% higher than with Universal ART) or 288,900–326,800 until 2050 (25% higher than with Universal ART). With Test&Treat, the ranges dropped to 46,200–49,400 until 2020 or 214,900–241,500 until 2050, about 7%–8% lower than with Universal ART.

The differences between monitoring and retention strategies were smaller (Figure 3, Table 3). Assuming the Universal ART scenario, current retention, no tracing, and no VL monitoring, 53,400 patients were infected between 2017 and 2020, or 260,900 between 2017 and 2050. Monitoring VL at 24-month intervals lowered these numbers by 0.5% to 53,200 until 2020, or by 0.9% to 258,700 until 2050; and, at 12-month intervals, by 0.8% to 53,000 until 2020, or by 1.4% to 257,200 until 2050. The relative differences remained the same across all scenarios of treatment scale-up and retention/tracing.

In the Universal ART scenario without viral load monitoring, actively tracing patients LTFU reduced the average number of new infections in 2017–2020 by 0.3% from 53,400 to 53,300, or in 2017–2050 by 0.5% from 260,900 to 259,500 (Figure 3, Table 3). When treatment interruptions were eliminated, the number of new infections decreased in 2017–2020 by 5.5% to 50,500, or in 2017–2050 by 9.9% to 235,100. The relative benefit of tracing and improved retention was similar across scenarios of treatment scale-up and monitoring.

The total number of PLHIV also decreased over time. In 2020, we predicted 839,200–849,500 PLHIV, and in 2050, 465,300–538,000 PLHIV. Because of the decreasing number of PLHIV, the relative differences remained the same across all scenarios of treatment scale-up and retention/tracing.

Table 2. Parameters of the Transmission Model: Parameters With Values Depending on Time Period and ART Initiation Scenario

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of starting ART, asymptomatic, adults</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.1</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>Rate of starting ART, symptomatic, adults</td>
<td>0.05</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rate of starting ART, asymptomatic, children</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Rate of starting ART, symptomatic, children</td>
<td>0.3</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Rate of starting ART due to PMTCT, asymptomatic, women</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Rate of starting ART due to PMTCT, symptomatic, women</td>
<td>0.15</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Rate of AIDS for patients on treatment</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

All rates are per person-year.

Abbreviations: ART, antiretroviral therapy; PMTCT, prevention of mother-to-child transmission.
Table 3. Number of Predicted HIV Infections in Malawi in Different Scenarios

<table>
<thead>
<tr>
<th></th>
<th>Failed Scale-up</th>
<th>Universal ART</th>
<th>Test&amp;Treat</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) CD4 monitoring, no tracing</td>
<td>13363</td>
<td>60 182 (ref)</td>
<td>326 790 (ref)</td>
</tr>
<tr>
<td>ii) CD4 monitoring, tracing</td>
<td>13314</td>
<td>60 023 (0.3%)</td>
<td>324 965 (0.6%)</td>
</tr>
<tr>
<td>iii) CD4 monitoring, no interruptions</td>
<td>12438</td>
<td>57 218 (4.9%)</td>
<td>293 353 (10.2%)</td>
</tr>
<tr>
<td>iv) 24-m VL monitoring, no tracing</td>
<td>13284</td>
<td>59 927 (0.4%)</td>
<td>323 868 (0.9%)</td>
</tr>
<tr>
<td>v) 24-m VL monitoring, tracing</td>
<td>13289</td>
<td>59 942 (0.4%)</td>
<td>324 039 (0.8%)</td>
</tr>
<tr>
<td>vi) 24-m VL monitoring, no interruptions</td>
<td>12384</td>
<td>57 044 (5.2%)</td>
<td>291 420 (10.8%)</td>
</tr>
<tr>
<td>vii) 12-m VL monitoring, no tracing</td>
<td>13232</td>
<td>59 760 (0.7%)</td>
<td>321 959 (1.5%)</td>
</tr>
<tr>
<td>viii) 12-m VL monitoring, tracing</td>
<td>13154</td>
<td>59 511 (1.1%)</td>
<td>319 122 (2.3%)</td>
</tr>
<tr>
<td>ix) 12-m VL monitoring, no interruptions</td>
<td>12312</td>
<td>56 815 (5.6%)</td>
<td>288 888 (11.6%)</td>
</tr>
</tbody>
</table>

Percentages in parentheses refer to the reduction compared with scenario (i) (first row) of the corresponding access scenario (Failed scale-up, Universal ART, or Test&Treat).

Abbreviations: ART, antiretroviral therapy; VL, viral load.

In this modeling study, we found that the number of new HIV infections will likely continue to decrease rapidly in all testing, treatment, retention, and monitoring scenarios. The most important factors associated with the speed of decrease were the rates of ART initiation and treatment interruptions (Supplementary Table 6). If we assumed lower mortality, the absolute number of new infections increased, but the relative benefit of the monitoring and retention support strategies did not differ from the main analysis (Supplementary Table 7).

Sensitivity Analyses

If the risk of virological failure increased over time, monitoring strategy no longer had an influence on the number of new infections (Supplementary Table 6). If we assumed lower mortality, the absolute number of new infections increased, but the relative benefit of the monitoring and retention support strategies did not differ from the main analysis (Supplementary Table 7).

DISCUSSION

In this modeling study, we found that the number of new HIV infections will likely continue to decrease rapidly in all testing, treatment, retention, and monitoring scenarios. The most important factors associated with the speed of decrease were the rates of ART initiation and treatment interruptions (Supplementary Table 6). If we assumed lower mortality, the absolute number of new infections increased, but the relative benefit of the monitoring and retention support strategies did not differ from the main analysis (Supplementary Table 7).

In this study, we predicted that routine VL monitoring could prevent up to one-third of transmissions from treated patients [12]. But the proportion of new infections attributable to patients on treatment with accelerated ART scale-up and faster ART scale-up lowered the number of patients on ART, which, in 2050, was 8% higher with Failed scale-up, and 2% lower with Test&Treat, than with Universal ART.

The results of VL monitoring were expected. In an earlier study, we predicted that routine VL monitoring could prevent up to one-third of transmissions from treated patients [12]. But the proportion of new infections attributable to patients on treatment with accelerated ART scale-up and faster ART scale-up lowered the number of patients on ART, which, in 2050, was 8% higher with Failed scale-up, and 2% lower with Test&Treat, than with Universal ART.
ART is low, so it is unsurprising that the overall effect of VL monitoring was small. Our results are in line with other modeling studies. For example, in 2014, Braithwaite et al. noted that VL monitoring was more cost-effective if its effect on transmission was considered, but extending ART eligibility, as the WHO recommended, had much more impact [33]. Some potential advantages we did not include in our model could also increase the benefit of VL monitoring. Patients whose ART is failing despite good adherence may believe they are not infectious, and thus be likely to engage in unprotected sex [34]. Patients on failing ART may also carry resistant strains of HIV. Their spread could limit the potency of available firstline regimens, increasing long-term mortality [35].

Loss to follow-up remains a major problem in sub-Saharan Africa. Although much LTTF can be explained by undocumented deaths and transfers between facilities, about one-third of lost patients have probably stopped treatment or are taking ART irregularly [16]. There is a broad range of reasons contributing to interrupting treatment [36]. Especially during the introduction of the Option B+, there were also concerns about retention rates among women who start ART in the asymptomatic stage. Experience from this program has, however, shown that long-term retention is feasible even among patients who start ART in an early stage of the infection [37]. In a prior study, we found that tracing lost patients did not substantially reduce expected transmission from patients who had started ART [13]. This study confirmed our finding. Tracing rarely locates patients who have moved or are traveling, or who provided an incorrect address or phone number. Many patients refuse to return to care, or interrupt treatment again later [16]. If treatment interruptions could be eliminated, the overall number of new infections would drop by 5% in the next few years, but perfect retention is unrealistic. Retention may be increased by further decentralizing treatment services, providing larger supplies of ART per visit, or SMS reminders [16, 38–40].

A recent survey found that, in Malawi, 89% of diagnosed PLHIV were on ART, but only 73% of all PLHIV had been diagnosed. Moreover, 91% of patients tested for VL in Malawi were virally suppressed [41]. Our results suggest that continuing screening to find more PLHIV in Malawi is the most effective strategy for meeting the ambitious 90-90-90 target. Our model predicted that, by 2020, in all scenarios, at least 90% of all PLHIV would be on ART. We may slightly overestimate the number of PLHIV on ART, as our estimate includes patients who have interrupted ART. But it is clear the 90-90-90 target can realistically be met in Malawi.

Cost-effectiveness calculations of VL monitoring and tracing patients LTTF should account for benefits to both individuals and the population. If VL test costs could be suppressed to $10 [10], testing each patient annually would cost about $6 million. This would prevent only about 100 new infections each year, thus costing about $60,000 to preventing a single infection. This is clearly more expensive than treating the infected patient for life. Annual VL testing may, however, also benefit the patient and reduce transmitted drug resistance. A more complete perspective is needed to assess cost-effectiveness.

Strengths and Limitations

We used a structurally simple, deterministic compartmental transmission model to produce results that closely match the data and projections returned by other established mathematical models. Our transmission model was informed by an individual-based simulation of the progression of HIV under ART, parameterized with routine cohort data. We expect that our results are generalizable to many Southern and Eastern African countries with similar epidemics and relatively high ART coverage [14, 41]. We also expect that the relative decrease in new infections with VL monitoring or tracing will be similar in other countries in this region. Our approach of linking a detailed simulation of disease progression to a relatively simple deterministic transmission model can also be adapted to other settings and questions.

Our study also had several limitations. The data we used to parameterize the model covered only a few years of follow-up and may thus not be accurate for long-term projections. We also did not consider other future changes in care, such as new antiretroviral regimens. New regimens may lead to substantially lower failure rates. If third-line ART becomes widely available in the future, it will likely further reduce the number of new infections, influencing also the relative benefit of the monitoring and retention interventions. Apart from a resistance penalty factor, first- and second-line failures were sampled independently, ignoring the possible individual factors that may result in failure through, for example, poor adherence. Several parameters had to be estimated using assumptions and calibration. We aimed to keep the transmission model as simple as possible and did not divide patients according to VL strata during treatment. Instead, we calculated average infectiousness for each stage of the disease, so our model may not be able to catch variation in infectiousness over time. We included only heterosexual transmission. Finally, we did not account for geographical variability, which can play a major role in the dynamics of the epidemic: district-level HIV prevalence in Malawi ranges from 3% in some northern districts to over 18% in the south.

There were also discrepancies between our and UNAIDS’ modeling estimates. In particular, our model estimated the annual new infections higher than UNAIDS. This resulted also in a slightly differing pattern in the total number of PLHIV. We predicted a moderate decrease since 2000, whereas UNAIDS suggests that the number remained stable or even increased slightly. However, the UNAIDS estimates are also based on models built on a limited amount of data. Our model’s
projections on number of patients on ART matched closely with the reported data. However, our estimate also includes patients who have stopped ART.

**Conclusions**

Malawi’s response to the HIV epidemic has been successful, and we expect the number of new infections to continue to fall, regardless of the strategies of ART initiation, monitoring, and retention support. To reach the 90-90-90 goal, scaling-up ART and preventing treatment interruptions must be a priority. Though only a small number of infections can be prevented by viral load monitoring and tracing patients lost to follow-up, the population-level effects of these interventions should be included in future cost-effectiveness evaluations.

The substantial heterogeneity in the Malawian HIV epidemic across regions and population groups may require interventions targeted to specific groups or regions to increase treatment efficiency and acceptability. Test&Treat reduces transmission of HIV, despite individual cases of treatment failure and ART interruption, but the full benefits of the treatment as a prevention strategy cannot be reaped unless ART is improved with interventions that support retention and viral suppression.

**Supplementary Data**

Supplementary materials are available at Open Forum Infectious Diseases online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the corresponding author.

**Acknowledgments**

We thank Kali Tal for commenting on and editing the manuscript.

**Financial support.** This work was supported by the Swiss National Science Foundation (grant numbers PP00P3_163878, PDPMP3_137106).

**Potential conflicts of interest.** All authors: no reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**

31. Wandel S, Egger M, Rangsin R, et al. Duration from seroconversion to eligibility of ART initiation, monitoring, and we expect the number of new infections to continue to fall, regardless of the strategies of ART initiation, monitoring, and retention support. To reach the 90-90-90 goal, scaling-up ART and preventing treatment interruptions must be a priority. Though only a small number of infections can be prevented by viral load monitoring and tracing patients lost to follow-up, the population-level effects of these interventions should be included in future cost-effectiveness evaluations.

The substantial heterogeneity in the Malawian HIV epidemic across regions and population groups may require interventions targeted to specific groups or regions to increase treatment efficiency and acceptability. Test&Treat reduces transmission of HIV, despite individual cases of treatment failure and ART interruption, but the full benefits of the treatment as a prevention strategy cannot be reaped unless ART is improved with interventions that support retention and viral suppression.


