# Viral load monitoring of antiretroviral therapy, cohort viral load and HIV transmission in Southern Africa: a mathematical modelling analysis

Janne Estill[a], Cindy Aubrière[a], Matthias Egger[a], Leigh Johnson[b], Robin Wood[c], Daniela Garone[d], Thomas Gsponer[a], Gilles Wandeler[a,f], Andrew Boulle[b], Mary-Ann Davies[b], Timothy B. Hallett[e], Olivia Keiser[a], for IeDEA Southern Africa

**Objectives:** In low-income settings, treatment failure is often identified using CD4 cell count monitoring. Consequently, patients remain on a failing regimen, resulting in a higher risk of transmission. We investigated the benefit of routine viral load monitoring for reducing HIV transmission.

**Design:** Mathematical model.

**Methods:** We developed a stochastic mathematical model representing the course of individual viral load, immunological response and survival in a cohort of 1000 HIV-infected patients receiving antiretroviral therapy (ART) in southern Africa. We calculated cohort viral load (CVL; sum of individual viral loads) and used a mathematical relationship between individual viral load values and transmission probability to estimate the number of new HIV infections. Our model was parameterized with data from the International epidemiologic Databases to Evaluate AIDS Southern African collaboration. Sensitivity analyses were performed to assess the validity of the results in a universal 'test and treat' scenario, wherein patients start ART earlier after HIV infection.

**Results:** If CD4 cell count alone was regularly monitored, the CVL was $2.6 \times 10^6$ copies/ml and the treated patients transmitted on average 6.3 infections each year. With routine viral load monitoring, both CVL and transmissions were reduced by 31% to $1.7 \times 10^6$ copies/ml and 4.3 transmissions, respectively. The relative reduction of 31% between monitoring strategies remained similar for different scenarios.

**Conclusion:** Although routine viral load monitoring enhances the preventive effect of ART, the provision of ART to everyone in need should remain the highest priority.



*AIDS* 2012, **26**:1403–1413

**Keywords: HIV transmission, mathematical model, second-line therapy, Southern Africa, test and treat, therapy failure, viral load monitoring**

[a]Institute of Social and Preventive Medicine, University of Bern, Switzerland, [b]Centre for Infectious Disease Epidemiology & Research, School of Public Health & Family Medicine, University of Cape Town, [c]Desmond Tutu HIV Centre, Institute for Infectious Disease & Molecular Medicine, University of Cape Town, [d]Khayelitsha ART Programme, Médecins Sans Frontières, Cape Town, South Africa, [e]Department of Infectious Disease Epidemiology, Imperial College London, London, UK, and [f]Infectious Diseases Clinic, University Hospital Bern, Switzerland.

Correspondence to Olivia Keiser, Institute of Social and Preventive Medicine (ISPM), University of Bern, Finkenhubelweg 11, CH-3012 Bern, Switzerland.

Tel: +41 31 631 35 15; fax: +41 31 631 35 20; e-mail: okeiser@ispm.unibe.ch
Received: 6 November 2011; revised: 17 February 2012; accepted: 7 March 2012.

DOI:10.1097/QAD.0b013e3283536988

## Introduction

In the year 2010, 2.7 million people were newly infected with HIV [1]. The majority of infections occur in sub-Saharan Africa, where resources for patient management remain limited. There is a strong association between plasma HIV viral load and the risk of HIV transmission [2]; meta-analyses of studies of serodiscordant couples showed that patients who were treated with antiretroviral therapy (ART) and had undetectable viral load did not transmit HIV [3,4]. A more recent randomized controlled trial showed that early initiation of ART reduced transmission of HIV [5]. 'Test and treat', which involves large-scale testing for HIV infection and immediate ART, is a subject of debate [6–8]. The acceptability and feasibility of universal testing and treatment is unclear, and modelling studies have yielded conflicting results [6,9,10]. In particular, it may be difficult to achieve the necessary high levels of adherence to therapy and high-risk sexual behaviours might increase.

Relatively little attention has been paid to the fact that with a 'test and treat' approach, not only the number of people on ART, but the number of individuals failing ART will increase. In high-income settings, viral load is measured regularly to detect treatment failure and counsel patients on adherence [11]. In sub-Saharan Africa and other low-income settings, viral load monitoring is not generally available and ART programmes, therefore, rely on immunological and clinical criteria to identify treatment failure [11]. The WHO CD4 cell count criteria are, however, inaccurate predictors of virological failure [12,13]. In Malawi and Zambia, where monitoring is based on CD4 cell counts, a study showed that few patients switched to second-line ART and many more remained on a failing first-line regimen compared to ART programmes monitoring viral load in South Africa [14]. Routine viral load monitoring may, thus, help to prevent new HIV infections by reducing the number of patients on failing first-line regimens.

We developed an individual-based mathematical model to study the importance of routine viral load monitoring versus CD4 cell monitoring on cohort viral load (CVL) and HIV transmission in Southern Africa. We analyzed data from two sites participating in a collaboration of HIV treatment programmes in Southern Africa to parameterize the model and the results of these analyses are also presented.

## Methods

### Data sources, eligibility and definitions

The International epidemiologic Databases to Evaluate AIDS in Southern Africa (IeDEA-SA) is a collaboration of ART programmes in six countries in Southern Africa [15]. Data are collected at ART initiation (baseline) and each follow-up visit using standardized instruments. All sites have ethical approval to collect data and participate in IeDEA-SA.

We restricted our analyses to the Gugulethu and Khayelitsha ART programmes in Cape Town, South Africa, where viral load and CD4 cell counts are measured regularly. All treatment-naive patients aged at least 16 years, who had started ART with at least two nucleoside reverse transcriptase inhibitors (NRTIs) and one non-nucleoside reverse transcriptase inhibitor (NNRTI) were included. Second-line ART was defined as a switch from an NNRTI-based regimen to a protease inhibitor-based regimen, with at least one NRTI changed.

We conducted statistical analyses on the cohort data and did literature searches to estimate parameters. We used parametric and semiparametric models to estimate time to virological failure (viral load over 1000 copies/ml), immunological failure (according to WHO criteria [16]), death or loss to follow-up. Observed mortality, loss to follow-up and non-HIV background mortality according to the ASSA2008 model for Africans in Western Cape in 2007 [17] were used to calculate the corrected estimate for HIV-related mortality. Details on the statistical and mathematical methods are given in the web appendix (1.1–1.2, http://links.lww.com/QAD/A213).

### Mathematical model

We adapted a published model that simulated disease progression in a hypothetical cohort of 1000 HIV-infected patients prior to starting ART [18]. In our model, individuals were simulated independently of each other and the properties of the individual and the timing of events were calculated probabilistically based on a series of rules and parametric distributions. The model included a description of the time period before start of ART and a detailed description of the time from ART start to either death or a fixed maximum follow-up time. In the following paragraph we give a brief explanation of the structure of the model; more details are found in the web appendix (3.1–3.2 http://links.lww.com/QAD/A213).

#### *Modelling of treatment response and mortality*

Each patient was assigned a baseline age and sex and, based on these, an HIV-free life expectancy was determined. Times of virological and immunological failure were defined by simulating from distributions parameterized by the data. Depending on the chosen monitoring strategy either viral load or CD4 cell count is measured every 6 months. Failures are observed at the next visit after the true unobserved failure, and confirmed in a second measurement 3 months later. Once failure is observed, the patient switches to second-line ART. Second-line failures are defined in the same way as first-line failures using parameters from data, but their

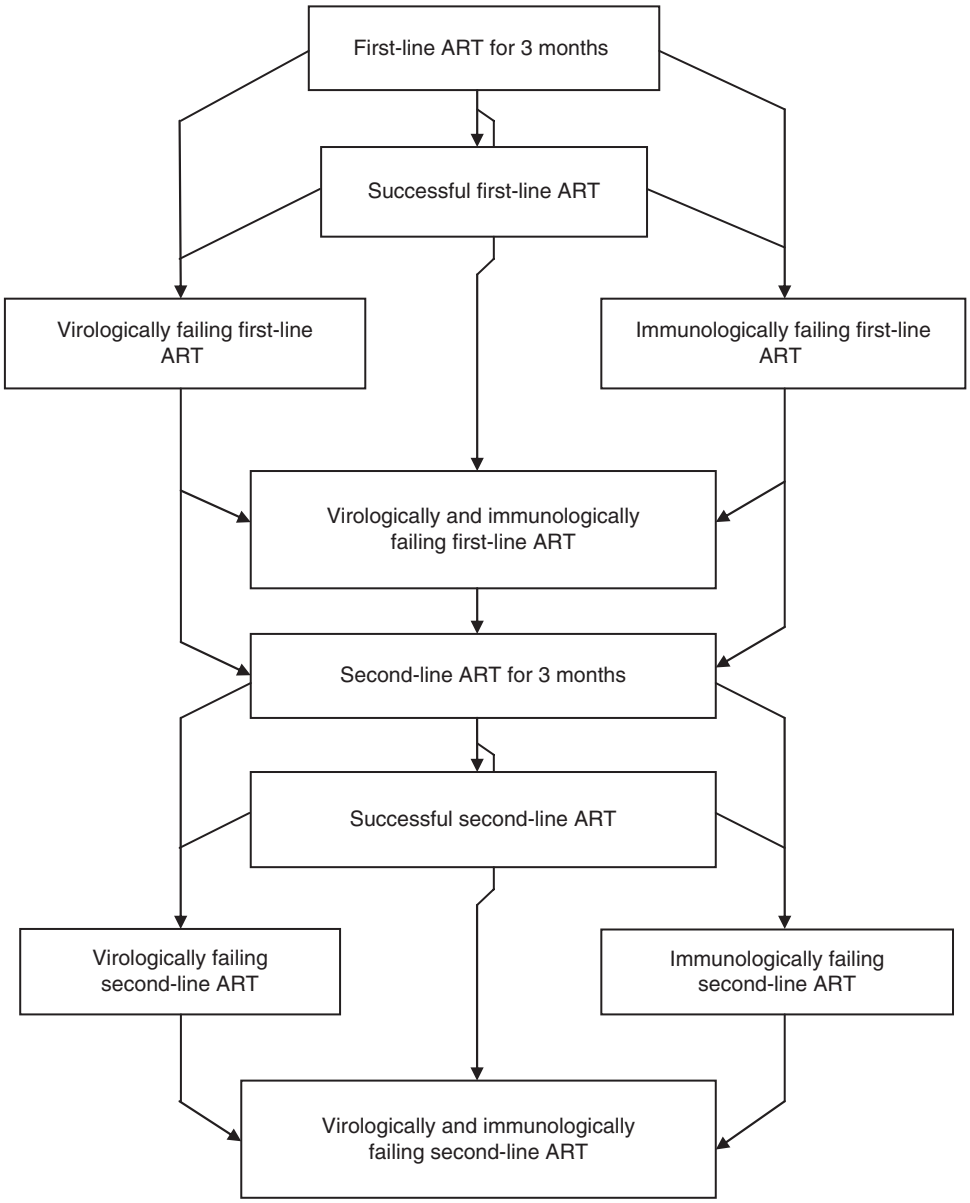


**Fig. 1. Flow chart showing possible transitions between risk stages after start of antiretroviral therapy.** In each stage, the patient is exposed to specific risk of death. Transitions between stages are assigned randomly according to the failure and switching rates observed in the Gugulethu and Khayelitsha antiretroviral therapy (ART) programmes in Cape Town, South Africa. The flow chart is simplified; the nature of immunological failure (with or without preceding virological failure) will influence outcome of second-line ART. As treatment failures are rare, most patients stay on successful first-line ART during the entire follow-up period.

probability additionally depends on the time spent on a virologically failing first-line therapy and the first-line immunological response.

Failure and switching events split the patient's follow-up time into different portions, as illustrated in Fig. 1. Each patient spends the first 3 months in the initial stage, after which she either enters a phase of successful treatment or she experiences immunological or virological failure. After virological failure, the hazard of death is assumed to increase over time [19]. Similarly, immunological failure increases the risk of death. After starting second-line therapy, the patient enters a 3-month period during which the hazard of HIV-related death returns to the level before failure, unless second-line failure occurs. The timing of failures and switching define the mortality hazard, which again determines the time of HIV-related death. To evaluate the level of viral load at the different stages of ART, we analyzed data from the two cohorts. The methods and results of these analyses are described in the web appendix (1.4, http://links.lww.com/QAD/A213).

*Cohort viral load and number of new infections*

We calculated two measures for potential transmission. We defined the CVL in a manner similar to the community viral load by Das *et al.* [20]. An explicit number of expected HIV transmissions were calculated according to a relationship between individual viral load values and infectiousness [2,21]. Both methods are presented in more detail in the web appendix (2.1–2.3, http://links.lww.com/QAD/A213).

*Main analysis*

We ran 1000 simulations for both monitoring strategies (CD4 monitoring and routine viral load monitoring) and used the point estimates of the statistical analyses as parameters. In both strategies, patients had measurements every 6 months. If failure was observed, another measurement was taken 3 months afterwards. We calculated annual CVL and transmission, from the last year before ART until the 5th year on ART. Mean values were calculated over the 5 years on ART, which were used to estimate the relative reduction in CVL and transmission for routine viral load monitoring compared to CD4 cell monitoring. The results were presented as mean values over the 1000 simulations with 95% confidence intervals.

*Sensitivity and uncertainty analyses*

We conducted a range of sensitivity analyses to explore the impact of our assumptions on the results (Table 1) [22]. In the first three analyses we varied the assumptions about the course of the individual viral load values over time. In two additional analyses we explored the consequences of earlier ART initiation, that is, we assumed lower early mortality rates and lower failure rates. In two final sensitivity analyses we assumed that the time spent on a failing first-line regimen would not affect the risk of second-line failure and we changed our assumptions about the effect of virological failure on mortality. To assess the impact of the variability of key parameter estimates on the results, we performed an uncertainty analysis, wherein we sampled key parameter values before each simulation using Latin Hypercube Sampling. Details of this analysis are presented in web appendix (4.2, http://links.lww.com/QAD/A213).

## Results

We describe the outcomes of the mathematical model including all sensitivity analyses, wherein hypothetical cohorts of 1000 patients were simulated with either routine viral load or CD4 monitoring to compare transmission. The baseline characteristics of the data are shown in the web appendix (1.3; Table S1, http://links.lww.com/QAD/A213). The results of the statistical analyses and parameters for the distributions of time to virological and immunological failure, time to switching to second-line ART, and time to death are shown in Table 2 [17,19,23]. The hazard ratios for

**Table 1. Key assumptions of main and sensitivity analyses.**

| | Main scenario | Sensitivity analysis |
|---|---|---|
| 1) Level of undetectable viral load | Undetectable viral load is 10 copies/ml | Undetectable viral load is 100 copies/ml |
| 2) Level of viral load after failure | The median viral load after failure is 10 000 copies/ml | The median viral load after failure is either 1000 or 100 000 copies/ml |
| 3) Time from undetectable viral load until reaching failure threshold | Viral load starts to increase 3 months before reaching failure threshold of 1000 copies/ml | Viral load starts to increase 1 month before reaching failure threshold 1000 copies/ml |
| 4) Mortality (HIV related) | Time to HIV related death is estimated from the cohorts; a double Weibull distribution (weighted sum of two Weibull distributions: one with decreasing, one with constant or increasing hazard) is used to reflect the high mortality early after ART start | Time to HIV related death is estimated from the cohorts, but the first component of the double Weibull distribution, which represents the high risk of death in the first months after ART initiation, is omitted |
| 5) Virological failure | Time to virological failure is estimated from the cohorts; a Weibull distribution is used | Time to virological failure is estimated from the cohorts, but the scale parameter of the Weibull distribution is increased to correspond to a 50% lower hazard than in the main analysis. This is in accordance with studies showing a lower risk of virologic failure in people starting ART earlier with higher CD4 cell counts [22] |
| S1) Resistance penalty for risk of second-line failure | The time from switching to second-line failure depends on the amount of time spent on a failing first-line ART regimen (i.e. assuming a resistance penalty) | No resistance penalty is included (i.e. risk of 2nd-line failure after switching is the same as the risk of 1st-line failure after ART start) |
| S2) Effect of virological failure on mortality | Hazard ratio of HIV-related mortality (virologically failing compared to successful ART) increases over time | Hazard ratio of HIV-related mortality (virologically failing compared to successful ART) is constant over time |

For results of analyses S1 and S2, see web appendix (4.1, http://links.lww.com/QAD/A213). The parameter values of the main analyses are shown in Table 2.

**Table 2. Model parameters and data sources.**

| Outcome | Source | Statistical model | Starting | Censoring | Value (95% CI) | Dimension | Risk |
|---|---|---|---|---|---|---|---|
| (1) Time to virological failure | | | | | | | |
| | Cohorts | Parametric Weibull | 3 months from ART start | Death, LTFU, switch to 2nd-line | 0.47 (0.43–0.50)<br>3.30 (2.77–3.95) | Shape<br>Scale (100 years) | 5.6% (1 year after ART start) |
| (2) Time to immunological failure | | | | | | | |
| (a) After virological failure | Cohorts | Parametric exponential | Virological failure | Death, LTFU, switch to 2nd-line | 12.72 (9.59–16.88) | Mean (years) | 7.6% (1 year after virol. failure) |
| (b) Before virological failure | Cohorts | Parametric Weibull | 3 months from ART start | Death, LTFU, switch to 2nd-line, virological failure | 0.22 (0.20–0.25)<br>5.46 (3.14–9.51) | Shape<br>Scale ($10^6$ years) | 3% (1 year after ART start) |
| (3) Time to death and LTFU | | | | | | | |
| (a) Background mortality, male | ASSA2008 [17] | Parametric Weibull | Age of 5 years | n/a | 6.36<br>69.89 | Shape<br>Scale (years) | 5.9% (at age of 50)[e] |
| (b) Background mortality, female | ASSA2008 [17] | Parametric Weibull | Age of 5 years | n/a | 7.09<br>73.60 | Shape<br>Scale (years) | 3.0% (at age of 50)[e] |
| (c) Observed mortality | Cohorts | No specific model (competing risk analysis) | ART start | Virological or immunological failure, switch to 2nd-line | [d] | | |
| (d) Observed LTFU | Cohorts | No specific model (competing risk analysis) | ART start | Virological or immunological failure, switch to 2nd-line | [d] | | |
| (e) Mortality among LTFU | Analysis 3d, Egger [23] | No specific model (theoretical calculation) | n/a | n/a | [d] | n/a | |
| (f) HIV-related mortality | Analyses 3a, 3b, 3c, 3e | Theoretical calculation, double Weibull[a] | ART start | n/a | 0.88 (0.88–0.90)<br>0.35 (0.32–0.39)<br>1.00 (1.00–1.00)<br>64.60 (54.52–76.55)<br>0.08 (0.08–0.08) | Shape 1<br>Scale 1 (years)<br>Shape 2<br>Scale 2 (years)<br>Weight (1st component) | 8.4% (1 year after ART start) |
| (g) Extra hazard after immunological failure | Cohorts | Cox regression | Immunological failure[b] | Switch to 2nd-line[c] | 1.75 (1.15–2.67) | HR, constant over time | n/a |
| (h) Extra hazard after virological failure | Petersen [19], cohorts | Cox regression[f] | Virological failure[b] | Switch to 2nd-line | 1.07 (0.91–1.26) | HR per each 3 months | n/a |

Distributions of times to event were assumed to be exponential, Weibull or double Weibull distributed, based on the cohort data. Cohort data are from the Khayelitsha and Gugulethu ART programmes in Cape Town, South Africa. ART, antiretroviral therapy; ASSA, Actuarial Society of South Africa; CI, confidence interval; HR, hazard ratio; LTFU, loss to follow-up; n/a, not applicable.
[a]Analysis was done by fitting the theoretically calculated cumulative incidence to a double Weibull model using the Matlab curve fitting tool.
[b]Analysis was started from ART start, but the hazard ratio is only applied from this time point onwards.
[c]Not applicable if failure happens according to the risk described in 2(b).
[d]The results are given as a cumulative incidence function (without closed parametric form).
[e]Excluding mortality before age of 5 (i.e. the risk of dying between the ages of 5 and 50).
[f]Estimated by comparing results of Petersen *et al.* [19] and our data analyses.

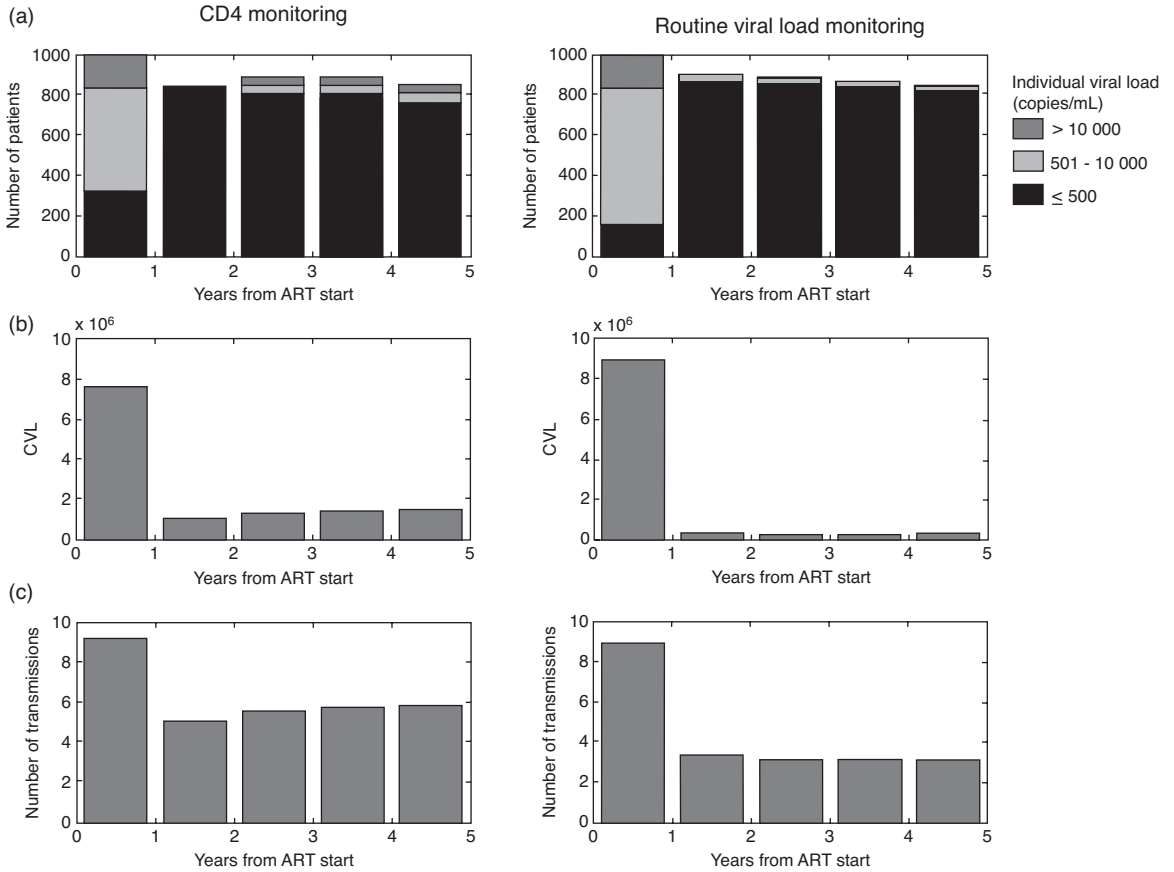


**Fig. 2. Number of patients in different viral load categories (a), cohort viral load (b) and expected number of new infections (c).** The same individuals were followed up through the entire 5 years, and because of mortality the total number of patients decreases over time. ART, antiretroviral therapy; CVL, cohort viral load.

mortality associated with virological and immunological failure are also shown.

## Cohort viral load and number of transmissions

The results during the first 5 years on ART are shown in Fig. 2. We assumed six-monthly CD4 monitoring alone (left panels) or routine viral load monitoring (right panels). The top panel (a) shows the number of patients alive at the beginning of each year in three viral load categories, panel (b) shows CVL, and panel (c) shows the expected number of new infections. In the last year before starting ART, CVL was $4.0 \times 10^8$ copies/ml, with most patients (99%) having viral load values above 10 000 copies/ml (results not shown). During the first year on ART, CVL was similar with both monitoring strategies ($7.6 \times 10^6$ copies/ml). During subsequent years, CVL ranged between $1.0 \times 10^6$ and $1.5 \times 10^6$ copies/ml with CD4 monitoring and between $2.3 \times 10^5$ and $2.9 \times 10^5$ copies/ml with routine viral load monitoring. The annual number of transmissions dropped from over 110 during the pre-ART period to approximately nine during the first year of ART. After this, the number of annual new infections ranged between three and four with routine viral load monitoring and five and six with CD4 monitoring. Routine viral load monitoring, therefore, reduced the number of new infections by about 30% (Table 3).

## Sensitivity and uncertainty analyses

Of the various sensitivity analyses (Table 3), the assumed level of the viral load value after failure (analysis 2) had the largest impact on the benefit of viral load monitoring relative to CD4 monitoring. Assuming 1000 copies/ml after failure resulted in a 5% reduction in CVL and a 16% reduction in number of new infections, relative to CD4 monitoring. If the viral load at failure was assumed to be similar to that at ART start, the corresponding reductions were 65 and 45%. Increasing the level of undetectable viral load to 100 copies/ml (analysis 1) led to an increase in CVL and the number of new infections. The benefit of routine viral load monitoring decreased; instead of 31 only 18% of infections were prevented, whereas the CVL decreased only slightly from 31 to 29%. The time that viral load started to increase before reaching the failure threshold value did not affect CVL or number of transmissions (analysis 3).

**Table 3. Cohort viral load and new infections in 1000 patients receiving ART over 5 years with either CD4 or routine viral load monitoring, under different assumptions.**

| | CD4 monitoring | | rVL monitoring | | Reduction (rVL compared to CD4) | |
|---|---|---|---|---|---|---|
| | CVL, $10^6$ copies/ml (95% CI) | Infections (95% CI) | CVL, $10^6$ copies/ml (95% CI) | Infections (95% CI) | CVL (95% CI) | Infections (95% CI) |
| Main scenario[a] | | | | | | |
| | 2.57 (2.05–3.09) | 6.25 (5.60–6.91) | 1.72 (1.38–2.05) | 4.33 (4.07–4.59) | 30.6% (11.3–49.9) | 30.6% (22.2–39.0) |
| Sensitivity analyses | | | | | | |
| 1) Level of undetectable viral load | | | | | | |
| 100 copies/ml | 2.81 (2.29–3.33) | 9.70 (9.06–10.35) | 1.97 (1.65–2.30) | 7.98 (7.71–8.26) | 29.9% (11.8–46.6) | 17.7% (11.6–23.7) |
| 2) Mean viral load after failure | | | | | | |
| 1000 copies/ml | 1.52 (1.26–1.78) | 4.59 (4.31–4.88) | 1.43 (1.17–1.69) | 3.86 (3.73–4.00) | 5.4% (−17.9–28.6) | 15.8% (10.0–21.6) |
| Equal to baseline (100 000 copies/ml) | 10.85 (7.11–14.58) | 9.53 (8.11–10.96) | 3.64 (2.43–4.84) | 5.17 (4.70–5.64) | 65.4% (49.0–81.9) | 45.4% (35.8–55.1) |
| 3) Mean time from undetectable viral load to reaching failure threshold (in patients experiencing virological failure | | | | | | |
| 1 month | 2.60 (2.08–3.13) | 6.27 (5.63–6.90) | 1.73 (1.43–2.03) | 4.33 (4.09–4.57) | 32.8% (15.4–50.3) | 30.7% (22.7–38.7) |
| 4) Mortality | | | | | | |
| Reduced | 2.68 (2.16–3.20) | 6.71 (6.04–7.39) | 1.73 (1.43–2.03) | 4.59 (4.36–4.83) | 34.8% (18.0–51.5) | 31.4% (23.4–39.2) |
| 5) Virological failure | | | | | | |
| Reduced | 1.98 (1.59–2.37) | 4.92 (4.45–5.39) | 1.53 (1.24–1.82) | 3.90 (3.73–4.07) | 21.9% (0.4–43.4) | 20.6% (12.4–28.7) |
| CD4 monitoring equal; rVL monitoring reduced | 2.57 (2.05–3.09) | 6.25 (5.60–6.91) | 1.53 (1.24–1.82) | 3.90 (3.73–4.07) | 39.6% (23.1–56.1) | 37.4% (30.4–44.4) |

CI, confidence interval; CVL, cohort viral load; rVL, routine viral load.
[a]Level of undetectable viral load = 10 copies/ml; mean viral load after failure = 10 000 copies/ml; mean time from undetectable viral load to reaching failure threshold = 3 months; risk of mortality: according to data; hazard of virological failure: according to data.

Lower mortality, which would be more realistic in a universal 'test and treat' strategy, did not affect CVL or number of transmissions (analysis 4). If failure rates were lower, both CVL and transmissions were reduced (analysis 5). If failure rates were assumed to be equally low in CD4 and viral load sites, 21% of transmissions were prevented and the reduction in CVL dropped to 22%. Assuming a lower failure rate in sites with routine viral load monitoring only, would prevent 37% of infections and lead to a 40% reduction in CVL.

The results of the two remaining sensitivity analyses (analyses S1 and S2) and the uncertainty analysis are shown in the web appendix (4.1–4.2, http://links.lww.com/QAD/A213).

## Discussion

There is an ongoing debate on the benefit of routine viral load monitoring. Regular viral load measurements help to detect treatment failure earlier and may, therefore, reduce mortality and HIV transmission. Our model shows that routine viral load monitoring reduces CVL substantially as compared to CD4 monitoring (currently the standard of care in many low-income countries [24,25]). Compared to CD4 monitoring, viral load measurements reduced the average CVL by more than 30% over 5 years, and the reduction in transmissions was similar.

At the end of 2009, ART coverage was below 40% in sub-Saharan Africa [26]. The time from HIV infection to ART eligibility is typically several years [27]; most people are tested too late for HIV [28] and the majority of the HIV-infected population remains untreated. The proportion of new infections from treated individuals is, therefore, small, and even large reductions in transmission from treated individuals would hardly reduce the viral load at the population level. However, there is a trend toward earlier ART initiation; recently, WHO increased the CD4 threshold for ART eligibility from 200 to 350 cells/μl [16], and there is evidence of clinical benefits for starting ART with CD4 values above 350 [29–31]. An even more fundamental change to the treatment policy would be the implementation of the 'test and treat' strategy [6–8].

Our results should also be valid in a 'test and treat' situation. Although they are based on patients starting ART with low CD4 values, the sensitivity analyses showed that similar benefits can be achieved when patients start ART earlier. Lower mortality did not have an impact on the benefit of routine viral load monitoring. Decreasing the rate of virological failure reduced the benefit of viral load monitoring, but this was offset by better adherence with viral load monitoring. Due to the current WHO guidelines, to our knowledge no data on mortality and virological failure rates have been published for people starting ART shortly after diagnosis of HIV. The available data do, however, suggest that the reduction in mortality due to higher CD4 cell count becomes minimal above 350 cells/μl [32], and several studies have found no association between baseline CD4 cell counts and risk of virological failure [33,34].

We investigated two different ways of estimating HIV transmission and both have advantages and disadvantages. Correlation between community viral load and HIV incidence has been described by Das *et al.* [20] in San Francisco. However, it is not clear if community viral load is a good proxy for HIV incidence in low-income settings, especially when assuming that most patients would be treated. The main advantages of CVL are that it is easy to calculate and it is independent of risk behaviour. One of the main limitations is that CVL does not take into account the number of individuals; 100 000 patients with undetectable viral load of 10 copies/ml, 1000 patients with a detectable viral load of 1000 copies/ml and one patient with a very high viral load of 1 000 000 copies/ml will all contribute the same amount to CVL, but the transmission potential will probably differ. The actual numbers of transmissions may, therefore, vary substantially between cohorts with similar CVL but different viral load distributions.

The other method we used to calculate transmissions assumes a linear relationship between log10 viral load values and HIV transmission. The resulting number of new infections is more intuitive than CVL. It can, for example, be directly transformed into costs, or other measures including the number needed to treat or the cost of preventing one HIV infection. However, calculating the absolute number of prevented transmissions is challenging, as it is highly sensitive to behavioural factors, which often are difficult to estimate.

Furthermore, the approximate reduction of 30% in transmissions should be applicable for different risk behaviour scenarios. In our calculations we assumed relatively high-risk behaviour. Individuals with fewer sexual acts and higher rate of condom use would transmit less, but the relative reduction would remain the same. The rate of partner change is also not a key factor, as the per-act transmission probabilities are low (see web appendix, 2.3 for more details, http://links.lww.com/QAD/A213).

Our model focused on a group of treated patients without considering the entire population, and this approach limited us in several ways. Because the main focus was on comparing monitoring strategies on ART, we did not model the pre-ART period in detail. Therefore, our results remain dependent on local characteristics, including HIV prevalence, ART coverage and ART eligibility criteria.

Transmission during the acute stage of infection has recently been estimated to contribute to about 38% of new infections [35] but various other estimates exist [36] and the acute phase may remain an important source of transmission in a 'test and treat' strategy. We did not take into account an increase in risk behaviour after ART start [37] and we did not consider behavioural differences that could result from different monitoring strategies apart from adherence. For example, patients from CD4 monitoring sites with unobserved virological failure and high CD4 cell counts may be more likely to engage in unprotected sex because they are unaware of the risk of transmitting the virus. If this was taken into account, probably even more infections could be prevented by viral load monitoring.

Similarly, we did not investigate the effect of possible worse adherence in patients starting ART with higher CD4 cell counts. The higher virological failure rates, which remain partly undetected in CD4 sites, would increase the benefit of viral load monitoring further. We also did not consider (primary) drug resistance that could complicate the treatment of newly infected individuals [38]. The assumption that the virologic failure rate, due to improved adherence counselling, would be 50% lower in viral load sites compared to sites without viral monitoring may appear high. It was an arbitrary choice that cannot be verified in our data. But a recent systematic review has shown that virologic failure rates vary substantially in sub-Saharan Africa [39] and they are also highly variable in viral load sites of the IeDEA-SA collaboration. Our results are only applicable for short term. To investigate the longer-term evolution of the epidemic, one would need a dynamic transmission model in which susceptible people get HIV infected and partnerships are modelled.

Our study was based on almost 10 000 adult patients from two public sector treatment programmes in South Africa. Results should, therefore, be applicable to many other patients in the region most heavily affected by HIV. We acknowledge that these treatment programmes will not be representative of all programmes in southern Africa; they are located in urban areas, are equipped with electronic medical record systems, and have access to regular CD4 cell determination, viral load monitoring and second-line therapy.

The availability of viral load monitoring may have led to an underestimation of immunologic failure rates, as patients should have switched after detection of virological failure. However, many patients never switched and the median time to switching from the estimated time of failure was 22 months. Moreover, limitations in the data required us to make assumptions about factors such as the effect of virological failure on mortality and the effect of the delay between failure and switching on second-line efficacy. However, in sensitivity analyses we found that these assumptions had little effect on the results (see web appendix, 4.1 for details, http://links.lww.com/QAD/A213).

The main barrier in providing routine viral load monitoring is its high cost. A recent randomized controlled trial estimated the difference in the unit cost between viral load and CD4 measurement to be approximately US$ 25. Therefore, the extra annual cost of treating 1000 patients with two viral loads instead of CD4 measurements per year would be about US$ 50 000. The net cost of preventing a new infection depends on the number of infections that can be prevented, as well as the total cost of treating and managing a new HIV-infected patient. In the United States, the discounted lifetime cost of a new HIV infection has been estimated to be over US$ 300 000 [40]. In low-income countries, these costs are much lower, for example, in Uganda, the total cost of treating a patient with ART and CD4 monitoring for a year is US$ 467. Assuming that patients spend on average at least 20 years on ART [41], the lifetime treatment costs would be around US$ 10 000. A detailed cost-effectiveness analysis is needed to evaluate whether routine viral load monitoring would be cost-effective or even cost saving in the long term. Such an analysis would, however, require more detailed information on sexual behaviour.

## Conclusion

After 15 years of ART and close to a decade of widespread ART use in low-income settings, it is still not clear if, when and how often viral load should be measured to optimize treatment outcomes. We found that viral load monitoring could be an important factor in reducing mortality [14], and could prevent HIV infections. Continuous evaluation of the role of routine viral load monitoring in terms of costs and effectiveness is necessary as new technologies are developed and new research findings become available. We emphasize that although the first priority should be providing ART, viral load monitoring could provide an additional benefit for ART as a preventive measure.

## Acknowledgements

We thank all study participants, Franziska Schöni-Affolter for helpful comments and Kali Tal for commenting and editing the manuscript.

J.E., O.K. and M.E. designed the study. R.W. and D.G. were involved in data acquisition and data management. J.E., C.A. and O.K. performed the statistical analyses. J.E., C.A. and T.H. developed the mathematical model. J.E. and O.K wrote the first draft of the manuscript. All authors contributed to the interpretation of the results and to the final version of the manuscript.

This study was supported by the National Institute of Allergy and Infectious Diseases (NIAID), Grant 5U01-AI069924-05, a PROSPER fellowship to O.K. supported by the Swiss National Science Foundation (Grant 32333B_131629) and a PhD student fellowship to J.E. from the Swiss School of Public Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Conflicts of interest

There are no conflicts of interest.

## References

1. World Heath Organization/UNAIDS/UNICEF. Global HIV/AIDS Response: Epidemic update and health sector progress towards Universal Access, (2011). http://www.who.int/hiv/pub/progress_report2011/hiv_full_report_2011.pdf. [Accessed 17 February 2011].
2. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, *et al*. **Viral load and heterosexual transmission of human immunodeficiency virus type 1.** *N Engl J Med* 2000; **342**:921–929.
3. Attia S, Egger M, Muller M, Zwahlen M, Low N. **Sexual transmission of HIV according to viral load and antiretroviral therapy: systematic review and meta-analysis.** *AIDS* 2009; **23**:1397–1404.
4. Anglemyer A, Rutherford GW, Egger M, Siegfried N. **Antiretroviral therapy for prevention of HIV transmission in HIV-discordant couples (Review)**. *Cochrane Database Syst Rev.* 2011:CD009153.
5. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, *et al*. **Prevention of HIV-1 infection with early antiretroviral therapy.** *N Engl J Med* 2011; **365**:493–505.
6. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. **Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model.** *Lancet* 2009; **373**:48–57.
7. Mastro TD, Cohen MS. **Honing the test-and-treat HIV strategy.** *Science* 2010; **328**:976.
8. Charlebois ED, Havlir DV. **"A Bird in the hand...": a commentary on the test and treat approach for HIV.** *Arch Intern Med* 2010; **170**:1354–1356.
9. Bacaer N, Pretorius C, Auvert B. **An age-structured model for the potential impact of generalized access to antiretrovirals on the South African HIV epidemic.** *Bull Math Biol* 2010; **72**:2180–2198.
10. Dodd PJ, Garnett GP, Hallett TB. **Examining the promise of HIV elimination by 'test and treat' in hyperendemic settings.** *AIDS* 2010; **24**:729–735.
11. Keiser O, Orrell C, Egger M, Wood R, Brinkhof MW, Furrer H, *et al*. **Public-health and individual approaches to antiretroviral therapy: township South Africa and Switzerland compared.** *PLoS Med* 2008; **5**:e148.
12. Keiser O, Macphail P, Boulle A, Wood R, Schechter M, Dabis F, *et al*. **Accuracy of WHO CD4 cell count criteria for virological failure of antiretroviral therapy.** *Trop Med Int Health* 2009; **14**:1220–1225.
13. Mee P, Fielding K, Charalambous S, Churchyard GJ, Grant AD. **Evaluation of the WHO criteria for antiretroviral treatment failure among adults in South Africa.** *AIDS* 2008; **22**:1971–1977.
14. Keiser O, Chi BH, Gsponer T, Boulle A, Orrell C, Phiri S, *et al*. **Outcomes of antiretroviral treatment in programmes with and without routine viral load monitoring in Southern Africa.** *AIDS* 2011; **25**:1761–1769.
15. Egger M, Ekouevi DK, Williams C, Lyamuya RE, Mukumbi H, Braitstein P, *et al*. **Cohort profile: the international epidemiological databases to evaluate AIDS (IeDEA) in sub-Saharan Africa**. *Int J Epidemiol* 2011 [Epub ahead of print].
16. World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach. 2010 revision. 2010. http://whqlibdoc.who.int/publications/2010/9789241599764_eng.pdf. [Accessed 17 February 2012].
17. Actuarial society of South Africa, (2011). http://www.actuarialsociety.org.za/. [Accessed 24 September 2011].
18. Hallett TB, Gregson S, Dube S, Garnett GP. **The impact of monitoring HIV patients prior to treatment in resource-poor settings: insights from mathematical modelling.** *PLoS Med* 2008; **5**:e53.
19. Petersen ML, van der Laan MJ, Napravnik S, Eron JJ, Moore RD, Deeks SG. **Long-term consequences of the delay between virologic failure of highly active antiretroviral therapy and regimen modification.** *AIDS* 2008; **22**:2097–2106.
20. Das M, Chu PL, Santos G-M, Scheer S, Vittinghoff E, McFarland W, *et al*. **Decreases in community viral load are accompanied by reductions in new HIV infections in San Francisco.** *PLoS One* 2010; **5**:e11068.
21. Wilson DP, Law MG, Grulich AE, Cooper DA, Kaldor JM. **Relation between HIV viral load and infectiousness: a model-based analysis.** *Lancet* 2008; **372**:314–320.
22. Fox MP, Sanne IM, Conradie F, Zeinecker J, Orrell C, Ive P, *et al*. **Initiating patients on antiretroviral therapy at CD4 cell counts above 200 cells/μl is associated with improved treatment outcomes in South Africa.** *AIDS* 2010; **24**:2041–2050.
23. Egger M, Spycher BD, Sidle J, Weigel R, Geng E, Fox MP, *et al*. **Correcting mortality for loss to follow up: a nomogram applied to antiretroviral treatment programmes in sub-Saharan Africa.** *PLoS Med* 2011; **8**:e1000390.
24. National Antiretroviral Treatment Guidelines for Adults, Adolescents, and Children. 3rd ed. Ministry of Health of Uganda, (2009). http://www.idi.ac.ug/docs/guidelines%202009.pdf. [Accessed 17 February 2012].
25. Johns Hopkins Point-of-Care Information Technology (POC-IT) Center. HIV Guide – Zambia. Zambia National HIV Guidelines. Baseline evaluation and monitoring, (2008). http://www.zambiahivguide.org/guidelines/index.html?categoryId=429187&siteId=429188. [Accessed 20 March 2012].
26. World Health Organization. Towards universal access: scaling up priority HIV/AIDS interventions in the health sector: Progress Report, (2010). http://www.who.int/hiv/pub/2010progressreport/full_report_en.pdf. [Accessed 17 February 2012].
27. Wandel S, Egger M, Rangsin R, Nelson KE, Costello C, Lewden C, *et al*. **Duration from seroconversion to eligibility for antiretroviral therapy and from ART eligibility to death in adult HIV-infected patients from low and middle-income countries: collaborative analysis of prospective studies.** *Sex Transm Infect* 2008; **84 (Suppl 1)**:i31–i36.
28. Bunnell R, Opio A, Musinguzi J, Kirungi W, Ekwaru P, Mishra V, *et al*. **HIV transmission risk behavior among HIV-infected adults in Uganda: results of a nationally representative survey.** *AIDS* 2008; **22**:617–624.
29. Kitahata MM, Gange SJ, Abraham AG, Merriman B, Saag MS, Justice AC, *et al*. **Effect of early versus deferred antiretroviral therapy for HIV on survival.** *N Engl J Med* 2009; **360**:1815–1826.
30. The HIV-CAUSAL Collaboration. **When to Initiate Combined Antiretroviral Therapy to Reduce Mortality and AIDS-Defining Illness in HIV-Infected Persons in Developed Countries: an observational study.** *Ann Intern Med* 2011; **154**:509–515.
31. Sterne JA, May M, Costagliola D, de Wolf F, Phillips AN, Harris R, *et al*. **Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies.** *Lancet* 2009; **373**:1352–1363.
32. Wright ST, Carr A, Woolley I, Giles M, Hoy J, Cooper DA, *et al*. **CD4 cell responses to combination antiretroviral therapy in patients starting therapy at high CD4 cell counts.** *J Acquir Immune Defic Syndr* 2011; **58**:72–79.
33. Lucas GM, Chaisson RE, Moore RD. **Highly active antiretroviral therapy in a large urban clinic: risk factors for virologic failure and adverse drug reactions.** *Ann Intern Med* 1999; **131**:81–87.
34. Phillips AN, Staszewski S, Weber R, Kirk O, Francioli P, Miller V, *et al*. **HIV viral load response to antiretroviral therapy according to the baseline CD4 cell count and viral load.** *JAMA* 2001; **286**:2560–2567.

35. Powers KA, Ghani AC, Miller WC, Hoffman IF, Pettifor AE, Kamanga G, *et al*. **The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study.** *The Lancet* 2011; **378**:256–268.
36. Cohen MS, Shaw GM, McMichael AJ, Haynes BF. **Medical progress: acute HIV-1 infection.** *N Engl J Med* 2011; **364**:1943–1954.
37. Bunnell R, Ekwaru JP, Solberg P, Wamai N, Bikaako-Kajura W, Were W, *et al*. **Changes in sexual behavior and risk of HIV transmission after antiretroviral therapy and prevention interventions in rural Uganda.** *AIDS* 2005; **20**:85–92.
38. Pingen M, Nijhuis M, de Bruijn JA, Boucher CAB, Wensing AMJ. **Evolutionary pathways of transmitted drug-resistant HIV-1.** *J Antimicrob Chemother* 2011; **66**:1467–1480.
39. Barth RE, van der Loeff MFS, Schuurman R, Hoepelman AIM, Wensing AMJ. **Virological follow-up of adult patients in antiretroviral treatment programmes in sub-Saharan Africa: a systematic review.** *Lancet Infect Dis* 2010; **10**:155–166.
40. Schackman BR, Gebo KA, Walensky RP, Losina E, Muccio T, Sax PE, *et al*. **The lifetime cost of current human immunodeficiency virus care in the United States.** *Med Care* 2006; **44**:990–997.
41. Mills EJ, Bakanda C, Birungi J, Chan K, Ford N, Cooper CL, *et al*. **Life expectancy of persons receiving combination antiretroviral therapy in low-income countries: a cohort analysis from Uganda.** *Ann Intern Med* 2011; **155**:209–216.