Supplemental Digital Content

Methods

1. Predicted change in patient CD4 and VL count

We carried forward a CD4+ cell (CD4) count or RNA HIV-1 (VL) count measurement for a maximum of 180 days if the next clinic visit measurement was missing, similar to Hoffmann et al., 2013.[1] We interpolated the remaining missing values between clinic visit dates no greater than twelve months using the ipolate function in Stata 12.1. Because of the wide range in test responses, we transformed all VL measurements onto the $\log_{10}$ scale. We obtained estimates for the mean change in CD4 and VL response for each patient using a fixed effects model. The model is given as

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, n_j;\ j = 1, \ldots, m \tag{1}$$

where for the *j*th patient and time point *i*: $\beta_{0j}$is the CD4/VL count at baseline (*i* = 1), β_{1j} is the monthly rate of change in CD4/VL count, x_{ij} represents the number of months between the test date and baseline date, and ϵ_{ij} is the *i*th statistical error.

We then computed a relative percentage change in patient CD4 count using the predicted values obtained from the fixed effects model at the end-date (*i* = 2) and six months prior to this date (*i* = 1). The predicted values were obtained with:

$$\tilde{y}_{ij} = \hat{\beta}_{0j} + \hat{\beta}_{1j} x_{ij}, \quad i = 1, 2;\ j = 1, \ldots, m \tag{2}$$

where for the *j*th patient: $\tilde{y}_{ij}$ is the point prediction (at time-point i = 1, 2), $\hat{\beta}_{0j}$ is the estimated intercept, $\hat{\beta}_{1j}$ is the estimated slope, x_{2j} is the period of observation, and x_{1j} is the period of observation less six months. The relative percentage change in a patient's predicted CD4 count over the last six months, abbreviated to %ΔCD4, was computed as follows:

$$\%\Delta\text{CD4}_j = \frac{\tilde{y}_{2j} - \tilde{y}_{1j}}{\tilde{y}_{1j}} \times 100. \tag{3}$$

We computed each patient's absolute change in predicted $\log_{10}$VL over the last six months, abbreviated to ΔVL, as follows:

$$\Delta\text{VL}_j = \tilde{y}_{2j} - \tilde{y}_{1j}. \tag{4}$$

A six month period was selected as a meaningful period for a clinician to assess a patient's medium term response to treatment and to diagnose a regimen switch to second-line ART. Using this approach, we acknowledge that the change in CD4 and VL response could be computed using the two most recent observed measurements. In this study, we justify the computation of patient-specific mean functions in order to a) include information from all patient CD4 and VL measurements during the period of observation, and b) to produce predicted values that are more robust to potentially large variability when only two observed measurements are used.

2. Classification of need for a second-line regimen switch

Having obtained the %ΔCD4 estimates, we classified each patient as having a high, medium, low or very-low need for a second-line regimen switch using the following step function:

$$\text{Need} = \begin{cases} \text{very-low} & \text{if } \ \%\Delta\text{CD4} > z \\ \text{low} & \text{if } \ y > \%\Delta\text{CD4} \leq z \\ \text{medium} & \text{if } \ x \geq \%\Delta\text{CD4} \leq y \\ \text{high} & \text{if } \ \%\Delta\text{CD4} < x, \end{cases} \qquad (5)$$

where $x < y < z$, and (x, y, z) are variables taking on real values called cut-points or thresholds. The need for a regimen switch using the ΔVL estimates was similarly determined:

$$\text{Need} = \begin{cases} \text{very-low} & \text{if } \ \Delta\text{VL} < x \\ \text{low} & \text{if } \ x \geq \Delta\text{VL} < y \\ \text{medium} & \text{if } \ y \geq \Delta\text{VL} < z \\ \text{high} & \text{if } \ \Delta\text{VL} \geq z, \end{cases} \qquad (6)$$

where $x < y < z$. We obtained the (x, y, z) cut-points for both monitoring strategies by using an algorithm to maximize the area under the curve (AUC) of a receiver operating characteristics (ROC) graph. The AUC, which ranges from zero to one, is an objective measure of the predictive performance of a diagnostic strategy. We show in eFigure 3 the ROC graph used to determine the CD4 (AUC=0.79) and VL (AUC=0.88) thresholds. The %ΔCD4 thresholds obtained with this method were x=20%, y=5%, and z=0%; and the ΔVL thresholds were x=−0.3, y=0.0, and z=0.3 $\log_{10}$ copies/ml.

As mentioned in the main text, we obtained measures for the sensitivity, specificity, false-positive rate (1 – specificity) and the positive predictive value (PPV). The inverse of the PPV reflects the number of patients that need to be tested (NNT) to make one correct second-line regimen switch. The best possible measure of the NNT is 1.0, with a higher measure indicating a less effective diagnostic strategy.[2]

In our analysis, we allow for a patient to have a discordant status, for example, having a high need for a regimen switch under immunologic monitoring, on the one hand, and a medium need for a regimen switch under virologic monitoring, on the other. This enables the diagnosis (need for a regimen switch) to individually and independently capture conflicting immunological and virologic responses in the same patient.[3]

3. Genotypic resistance testing

The true need for a regimen switch to second-line ART was determined by genotypic resistance testing. Patients in our sample with their latest two VL >1000 copies/ml were identified by clinic staff during routine visits (or proactively identified and contacted by staff) and referred to a physician for review. A 5 ml blood sample was collected during the clinical evaluation and an in-house HIV-1 drug resistance genotyping method was performed as previously described.[4] A genotypic susceptibility score (GSS) for each antiretroviral agent in the first-line regimen was determined using a Rega 8.0.0.2 algorithm, with a total GSS <2 indicating drug resistance. The outcome of this study was defined as a drug resistance result with the true need for a second-line regimen switch (n=396) or drug susceptibility on the first-line regimen (n=3781).

4. Survival curves and Cox proportional hazards model

Survival time was defined in months from the baseline measurement date until right censorship. The baseline date was defined as the most recent CD4/VL measurement prior to the date of ART initiation. Right censorship was defined as the date of the genotype test, or the last clinic date for those patients not sent for a genotype test. In Figure 1 of the main text we plot the time to drug

resistance for each need diagnoses (high, medium, low or very-low) using a Kaplan-Meier survival curve.

We also used a Cox proportional hazards (PH) analysis to model the time to drug resistance conditional on the need for a second-line regimen switch. We then added the covariates age and sex to the analysis. We did not adjust for duration on ART treatment since the interval between the start of survival time (date of the baseline test measurement) and the date of ART initiation was a median (IQR) of only 1.1 (0.5−1.9) months. When compared with a very-low need for a regimen switch, the unadjusted HR of drug resistance for medium need patients was increased by a factor of 2.7 (95% *CI*, 1.2 to 6.2; *P*=0.017) and by a factor of 11.8 (95% *CI*, 5.2 to 26.9; *P* <0.001) for high need patients under immunologic monitoring. When compared with very-low need patients, the unadjusted HR of drug resistance for medium need patients was increased by a factor of 38.9 (95% *CI*, 24.9 to 60.7; *P* <0.001) and by a factor of 92.9 (95% *CI*, 58.9 to 146.5; *P* <0.001) for high need patients under virologic monitoring. See eTable 1 for the HRs of drug resistance by need of a regimen switch adjusting for sex and age.

5. Cost-effectiveness model

We evaluated cost-effectiveness as a dollar amount for each threshold and monitoring strategy. This is the sum of 1) a baseline cost to make one correct regimen switch, and 2) a false-positive cost associated with incorrectly switching a patient to a more expensive second-line regimen for the duration of one year. We used the NNT to derive the baseline cost, under the assumption that a minimum of two measurements are needed to compute the change in CD4 or VL count over time. The baseline cost can be represented as:

$$\alpha = 2(NNT \times c), \qquad (7)$$

where c is the cost for a single CD4 ($9.18) or VL ($45.88) test at primary health care clinics in South Africa. It is likely that the cost of VL testing will decline in the near future. A WHO document reports that the cost of VL test is already US$15 for some low-income countries, although the authors indicate that this price varies considerably.[5] In this paper we consider the US$45 cost of a VL test so as to obtain a more conservative cost-savings estimate.

The cost associated with incorrectly switching patients to a second-line regimen is determined by a) the false-positive rate (fp), which is calculated as (1 − specificity), where the specificity is expressed as a proportion. The NNT is then multiplied by this false-positive rate to determine the proportion of patients that would be incorrectly switched to a more expensive second-line regimen for the duration of one year. The cost of a first-line regimen is $146.50 and a second-line regimen $465.50 for the duration of one year at public health-care clinics in South Africa. This false-positive cost can be expressed as:

$$\eta = fp \times NNT \times \delta, \qquad (8)$$

which, when added to α in equation (7), gives the cost to detect one correct regimen switch for each threshold and monitoring strategy. Let Λ be the total dollar amount to detect one correct regimen switch, then:

$$\Lambda = 2(NNT \times c) + (fp \times NNT \times \delta). \qquad (9)$$

We briefly outline a number of points with respect to our costing model presented in equation (9). We derive the cost-effectiveness of a treatment monitoring strategy as a function of its diagnostic performance. As a result, our model does not explicitly consider the logistical costs associated with implementing VL laboratory equipment and facilities, clinic visits, staff training, health-care worker time and other miscellaneous resources. However, additional costs scaled to either a single CD4 or VL test could feasibly be incorporated under the c component of equation (7). The cost to

correctly switch patients to a second-line regimen is a constant ($k \times \delta$) for both strategies, and is therefore not factored into the final costing equation. In addition, the total dollar amount does not include the cost of a patient's first-line regimen, which is also a constant under each monitoring strategy. While we include the cost of incorrectly switching patients to more expensive second-line ART for the duration of one year, it is possible to scale our model to consider the cumulative cost over a longer period adjusting for price inflation.

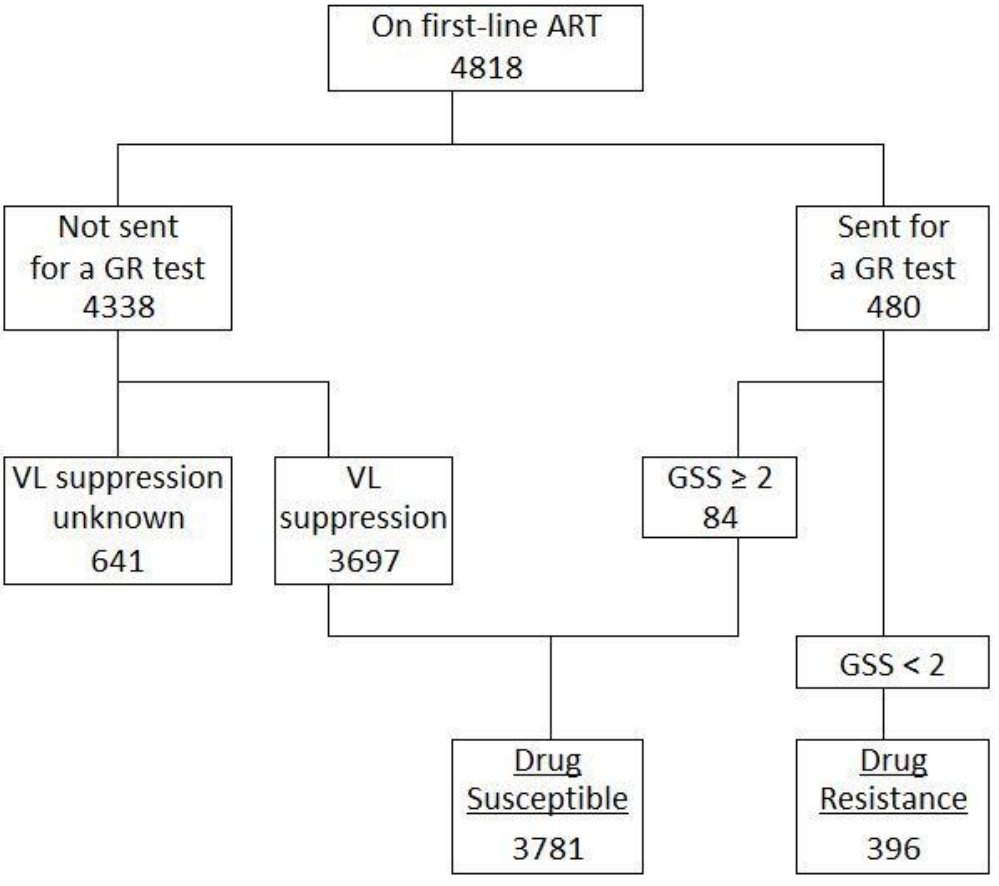
It should be noted that the frequency of testing will be dependent on current WHO or national guidelines. WHO guidelines now recommend routine VL testing six months after ART initiation and then every twelve months; and that CD4 be checked initially and at 12 months after ART initiation.[5, 6] Where routine VL monitoring is not available, WHO guidelines generally recommend CD4 testing every six months.[5] For this analysis, patients were scheduled for a CD4 test every six months, with VL tests scheduled at months six and twelve, and then every twelve months if VL <400 copies/ml or repeated after three months if VL >1000 copies/ml.[6]

Our costing model can be contrasted with quality-adjusted life years (QALYs), disability-adjusted life years (DALYs), or life years saved (LYS) studies, which are used predominantly within the economic evaluation framework to study the cost-effectiveness of immunologic and virologic monitoring.[7-11] These studies typically estimate the full health benefit of treatment monitoring (e.g., reduced mortality and HIV transmission, improved morbidity and ART adherence) using country- or region-specific weights for life expectancy, age, future time and disability. In some cases, assumptions about the dynamic transition of the population are factored into the health benefit projections of up to ten years (or longer).[12, 13] We preferred to focus our study on the clinical setting to reflect the balance (or trade-off) that has to be made between the predictive accuracy and the monetary cost of a treatment monitoring strategy. To conserve resources, a monitoring strategy should aim to detect treatment failure with the need for a regimen switch early and with minimal

cost, while ensuring that patients with virologic suppression are correctly identified so that they are not unnecessarily switched to a more expensive second-line ART regimen. We wanted to model this resource allocation problem from the particular perspective of the health-care strategist or clinician, whose efforts to optimize the detection of treatment failure is likely to be limited by budgetary constraints and the unavailability of drug resistance testing.

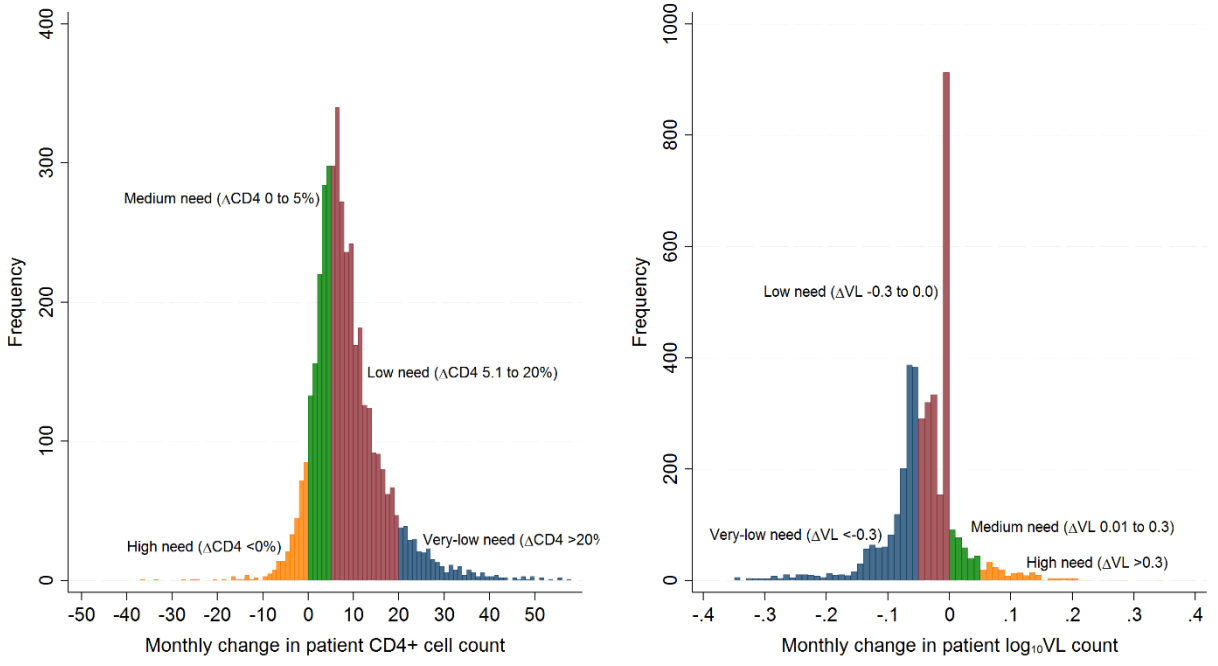
We acknowledge that our study is limited by the exclusion of 641 (11.45%) patients whose virologic failure status could not be definitively determined. To assess potential bias associated with this limitation, we assumed that these patients had drug resistance and included them in an alternative analysis. Results from this analysis do not deviate from the findings reported in this paper, where, for example, virologic monitoring is still proportionally more affordable than immunologic monitoring to make one correct regimen switch in high *and* medium need patients, and so on.

Figure S1: Data flow diagram of patients included and excluded from the study.



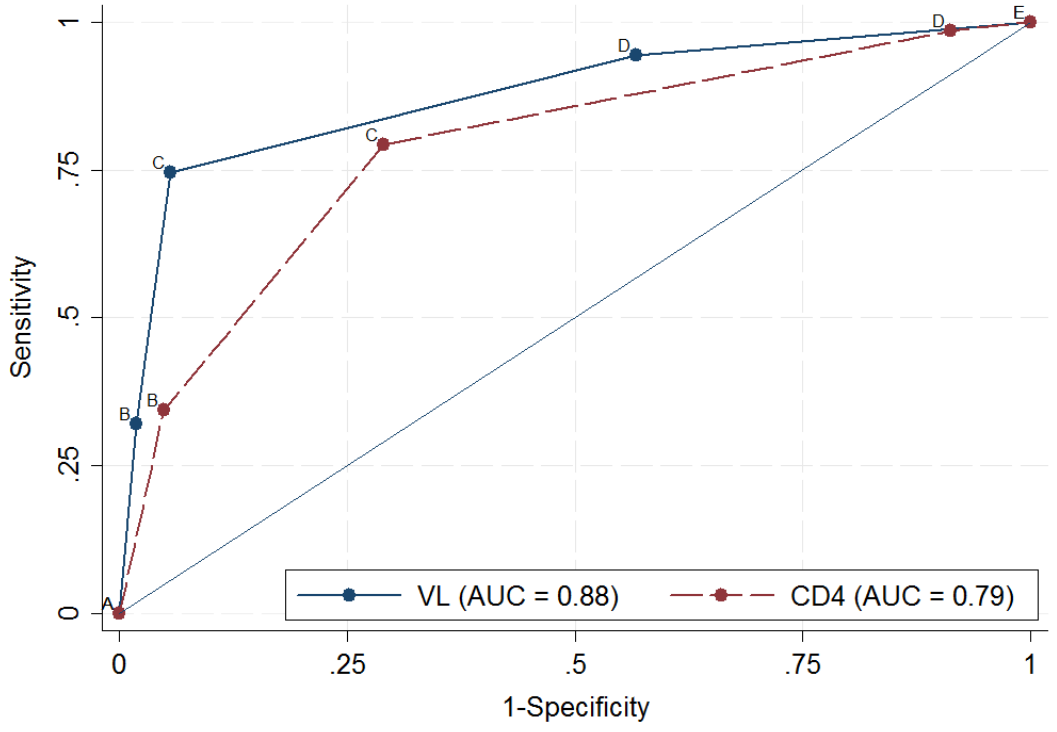
Abbreviations: VL, viral load; GR, genotypic resistance; GSS, genosusceptibility score. We included all patients (≥18 years) that were enrolled into the Hlabisa program between January 2006 and March 2014, who had two or more CD4/VL measurements, and were on a first-line ART regimen for >6months. Patients with the last two VL <400 copies/ml (n=3308) or undetectable VL <40 copies/ml at last clinic (n=389) were identified to have virologic suppression (n=3697). We excluded patients from the study whose treatment failure status could not be definitively determined (n=641).

Figure S2: Distributions of the patient-specific CD4 and $\log_{10}$VL slopes obtained from the fixed effects analysis.



The figure shows the distribution of the patient slopes for the percentage change in CD4 count (%ΔCD4, left panel) and absolute change in $\log_{10}$ viral load (ΔVL, right panel). We used the slopes to construct a qualitative measure of need for a second-line regimen switch by using the area under the curve (AUC) of a receiver operating characteristics (ROC) graph. Patients with a ΔCD4 <0% were described as having a high need for a regimen switch, ΔCD4 0.1 to 5.0% as having a medium need, ΔCD4 5.1 to 20.0% as having a low need and ΔCD4 >20.0% as having a very-low need. Similarly, we used VL thresholds of >0.3, 0.01 to 0.3, −0.3 to 0.0 and <−0.3 $\log_{10}$ copies/ml to respectively classify a high, medium, low and very-low need for a second-line regimen switch.

Figure S3: Receiver operator characteristic (ROC) graph showing the CD4 and VL thresholds used to classify the need for a second-line regimen switch.



The area under the curve (AUC) of a receiver operating characteristics (ROC) graph was maximized to obtain the ΔCD4 (0.0%, 5.0%, and 20.0% cells/μl) and ΔVL (0.3, 0.0, and −0.3 $\log_{10}$ copies/ml) thresholds labelled A, B, C, D, E. We used these thresholds to create a qualitative measure of the need for a second-line regimen switch. The five points on the graph show the sensitivity and false-positive rate (1 − specificity) for each of the following thresholds. A: no patients need a regimen switch; B: regimen switch in only high need patients; C: regimen switch in high and medium need patients; D: regimen switch in high, medium and low need patients; E: all patients need to be switched.

Table S1: Cox proportional hazards model showing the unadjusted and adjusted hazard ratios (HR) of drug resistance conditional on the need for a second-line regimen switch.

	Univariate			Multivariate		
	HR	**(95% CI)**	**p-value**	**HR**	**(95% CI)**	**p-value**
CD4 need group:						
Very-low	Ref.			Ref.		
Low	0.66	(0.28–1.51)	0.324	0.85	(0.37–1.96)	0.696
Medium	2.72	(1.20–6.17)	0.017	3.59	(1.58–8.18)	0.002
High	11.83	(5.21–26.88)	<0.001	13.59	(5.97–30.93)	<0.001
Age*				0.60	(0.53–0.67)	<0.001
Male				1.14	(0.92–1.42)	0.228
VL need group:						
Very-low	Ref.			Ref.		
Low	2.33	(1.45–3.75)	<0.001	2.33	(1.45–3.75)	0.001
Medium	38.89	(24.93–60.66)	<0.001	35.61	(22.81–55.58)	<0.001
High	92.92	(58.95–146.45)	<0.001	84.18	(53.34–132.86)	<0.001
Age*				0.68	(0.61–0.76)	<0.001
Male				1.24	(0.99–1.54)	0.057
N	4177			4177		

Time to treatment failure was calculated from the most recent CD4/VL test date prior to ART initiation until right censorship. The percentage change in absolute CD4 count (%ΔCD4) and absolute change in $\log_{10}$ viral load (ΔVL) was used to diagnose each patient's need for a regimen switch. This was diagnosed as either high (ΔCD4 <0%; ΔVL >0.3), medium (ΔCD4 0 to 5%; ΔVL 0.01 to 0.3), low (ΔCD4 5.1 to 20%; ΔVL -0.3 to 0) or very-low (ΔCD4 >20%; ΔVL <-0.3). The table shows that a higher hazard of drug resistance is more likely to be associated with a higher need diagnosis. For example, the unadjusted hazard of drug resistance is increased by a factor of 92.9 for high need patients when compared with very-low need patients under virologic monitoring. The results give an indication of how well the diagnoses (of need for a regimen switch) are able to correctly identify drug resistance in clinical contexts where resistance testing is either unavailable or too costly to undertake. *For a 10 year increase in age.

Table S2: Confusion matrix showing the diagnostic performance of immunologic and virologic monitoring.

Regimen Switch Need	Very-low		Low		Medium		High		Total	
Monitoring Strategy	**CD4**	**VL**	**CD4**	**VL**	**CD4**	**VL**	**CD4**	**VL**	**CD4**	**VL**
Drug resistance N	6	22	76	79	178	168	136	127	396	
No drug resistance N	333	1635	2352	1933	913	144	183	69	3781	
Total N	339	1657	2428	2012	1091	312	319	196	4177	

We created a qualitative measure of need for a second-line regimen switch based on each patient's change in CD4 (ΔCD4) and $\log_{10}$VL (ΔVL) count over the last six months. A patient's need was described as either high (ΔCD4 <0%; ΔVL >0.3), medium (ΔCD4 0 to 5%; ΔVL 0.01 to 0.3), low (ΔCD4 5.1 to 20%; ΔVL -0.3 to 0) or very-low (ΔCD4 >20%; ΔVL <-0.3).

We wanted to evaluate how accurately this qualitative measure of need (high, medium, low or very-low) could predict the *true* need for a second-line regimen switch. For example, the CD4 high need classifier correctly predicts 136 out of the 396 drug resistance cases, giving a specificity of 34.3%, and 3598 cases below this threshold are correctly classified out of the 3781 to have a drug susceptible profile, giving a specificity of 95.1%

References

1. Hoffmann CJ, Schomaker M, Fox MP, Mutevedzi P, Giddy J, Prozesky H, et al. CD4 Count Slope and Mortality in HIV-Infected Patients on Antiretroviral Therapy: Multicohort Analysis From South Africa. J Acquir Immune Defic Syndr. 2013;63(1):34-41.
2. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. The New England journal of medicine. 1988;318(26):1728.
3. Reynolds SJ, Sempa JB, Kiragga AN, Newell K, Nakigozi G, Galiwango R, et al. Is CD4 monitoring needed among ugandan clients achieving a virologic and immunologic response to treatment? AIDS patient care and STDs. 2014;28(11):575-8.
4. Manasa J, Lessells RJ, Skingsley A, Naidu KK, Newell M-L, McGrath N, et al. High-levels of acquired drug resistance in adult patients failing first-line antiretroviral therapy in a rural HIV treatment programme in KwaZulu-Natal, South Africa. PloS One. 2013;8(8):e72152.
5. WHO. March 2014 supplement to the 2013 consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach. 2014.
6. Who. Consolidated guidelines on the use of Antiretroviral drugs for treating and Preventing HIV infection: Recommendations for a public health approach: World Health Organization; 2013.
7. Keebler D, Revill P, Braithwaite S, Phillips A, Blaser N, Borquez A, et al. Cost-effectiveness of different strategies to monitor adults on antiretroviral treatment: a combined analysis of three mathematical models. The Lancet Global Health. 2014;2(1):e35-e43.
8. Estill J, Egger M, Blaser N, Vizcaya LS, Garone D, Wood R, et al. Cost-effectiveness of point-of-care viral load monitoring of ART in resource-limited settings: Mathematical modelling study. {AIDS}. 2013;27(9).
9. Boyer S, March L, Kouanfack C, Laborde-Balen Gel, Marino P, Aghokeng AF, et al. Monitoring of HIV viral load, CD4 cell count, and clinical assessment versus clinical monitoring alone for antiretroviral therapy in low-resource settings (Stratall ANRS 12110/ESTHER): a cost-effectiveness analysis. Lancet Inf Dis. 2013;13(7):577-86.
10. Phillips AN, Pillay D, Miners AH, Bennett DE, Gilks CF, Lundgren JD. Outcomes from monitoring of patients on antiretroviral therapy in resource-limited settings with viral load, CD4 cell count, or clinical observation alone: a computer simulation model. Lancet. 2008;371(9622):1443-51.
11. Braithwaite RS, Nucifora KA, Toohey C, Kessler J, Uhler LM, Mentor SM, et al. How do different eligibility guidelines for antiretroviral therapy affect the cost--effectiveness of routine viral load testing in sub-Saharan Africa? {{AIDS}}. 2014;28:S73-S83.
12. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health policy and planning. 2006;21(5):402-8.
13. Rushby JF, Hanson K. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. Health policy and planning. 2001;16(3):326-31.