Validation of a Clinical Prediction Score to Target Viral Load Testing in Adults With Suspected First-Line Treatment Failure in Resource-Constrained Settings

Vichet Phan, MD, Sopheak Thai, MD,* Olivier Koole, MD,† Joris Menten, MSc,† Filip Meheus, MSc,‡ Johan van Griensven, MD, PhD,*† and Lutgarde Lynen, MD, PhD†*

Background: Although routine viral load (VL) monitoring currently is too costly for poor countries, clinical failure criteria perform poorly. We previously developed an algorithm combining a clinical predictor score (CPS) with targeted VL testing in a Cambodian patient population (derivation population). We now prospectively validate the algorithm in the same clinical setting (validation population), assess its operational performance, and explore its cost-saving potential.

Methods: We performed a cross-sectional study in a tertiary hospital in Phnom Penh, Cambodia, applying the CPS in adults on first-line antiretroviral treatment for at least 1 year. Treatment failure was defined as a VL >1000 copies per milliliter. The area under the receiver-operating characteristic (AUROC) curve of the CPS to detect treatment failure in the current study population (validation population) was compared with the AUROC of the CPS obtained in the patient population where the CPS was derived from in 2008 in the same study setting (derivation population). Costs related to VL testing and second-line regimens with the different testing strategies were compared.

Results: One thousand four hundred ninety individuals {56.6% female, median age 38 years [interquartile range (IQR): 33–44]} were included, with a median baseline CD4 cell count of 94 cells per microliter (IQR: 28–205). Median time on antiretroviral treatment was 3.6 years (IQR: 2.1–5.1), 45 (3.0%) individuals had treatment failure. The AUROC of the CPS in validation was 0.75 (95% confidence interval: 0.67 to 0.83), relative to an AUROC of 0.70 in the derivation population. At the CPS cutoff ≥2, VL was indicated for 164 (11%) individuals, preventing inappropriate switching to second line in 143 cases. Twenty-four cases of treatment failure would be missed. When applied in routine care, the AUROC was 0.69 (95% confidence interval: 0.60 to 0.77). Overall 1-year program costs with targeted VL testing were 4-fold reduced.

Conclusions: The algorithm performed well in validation and has cost-saving potential. Further studies to assess its performance, feasibility, and impact in different settings are warranted.

Key Words: validation, algorithm, prediction score, HIV, viral load, antiretroviral, failure

(*J Acquir Immune Defic Syndr* 2013;62:509–516)

INTRODUCTION

Over the last couple of years, an impressive scaling up of antiretroviral treatment (ART) has taken place in low-income and middle-income countries (LMIC), with 6,650,000 people on treatment by the end of 2010.[1] One of the key but unresolved question is how individuals should be monitored for treatment failure in these settings.[2] In high-income countries, viral load (VL) testing is performed at regular intervals in routine care. Whether this should also be applied in LMIC is currently unclear. Several randomized clinical trials have failed to demonstrate substantial clinical benefit of including or adding routine VL monitoring in these settings.[3–5] Given its high cost, routine VL testing has not been considered a cost-effective use of scarce resources.[4,5] In line with World Health Organization (WHO) recommendations, most programs have been applying the WHO clinical and immunological failure criteria.[6] However, studies have consistently demonstrated the low sensitivity and specificity of these criteria, with positive predictive values ranging between 12% and 21%, demanding confirmation of failure by VL testing, to avoid unnecessary switch to second-line treatment.[7–9]

We previously developed a clinical prediction score (CPS) to detect treatment failure based on combined clinical adherence and laboratory data (Fig. 1).[7] The CPS was derived from an adult patient population treated at the Sihanouk Hospital Center of HOPE (SHCH) in Cambodia, who were on first-line ART since at least 1 year (derivation population). Within the derivation population, the score performed substantially better than the WHO failure criteria. Subsequently, the CPS was incorporated in a decision tree based on 3 probability levels of failure. A score <2 indicates low probability of failure and therefore, first-line treatment can be continued. A score of 2–4 indicates an intermediate risk of failure, but a VL is needed to confirm virological failure. A score of >4 indicates high probability of failure, and no VL is needed to switch to second-line ART.[7] We hypothesized this would be

Received for publication July 17, 2012; accepted December 28, 2012.

From the *Department of Infectious Diseases, Sihanouk Hospital Center of HOPE, Phnom Penh, Cambodia; and the Departments of †Clinical Sciences and ‡Public Health, Institute of Tropical Medicine, Antwerp, Belgium.

The HIV program was supported by the Belgian Directorate General of Development Cooperation through the framework agreement with the Institute of Tropical Medicine, Antwerp, Belgium; The Global Fund to fight AIDS, Tuberculosis, and Malaria; and Hope World Wide. J.V.G. is supported by the InBev-Baillet Latour Fund.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

The authors have no conflicts of interest to declare.

Correspondence to: Lutgarde Lynen, MD, PhD, Institute of Tropical Medicine, Nationale straat 155, 2000 Antwerp, Belgium (e-mail: llynen@itg.be).

Predictor	Score If Condition is Present
Adherence < 95% (using visual analogue scale)	3
CD4 count falls below baseline	1
> 25% CD4 count decrease from on-treatment peak	1
> 50% CD4 count decrease from on-treatment peak	1
CD4 count < 100 cells/μL after 12 months of ART	1
Prior exposure to HAART	1
Hemoglobin drop ≥ 1 g/dL	1
New onset or recurrent papular pruritic eruption	1
Maximum SCORE (Sum)	**10**

FIGURE 1. CPS to identify patient with possible first-line treatment failure. HAART, highly active antiretroviral therapy.

a cost-effective use of VL testing, with a primary aim of avoiding unnecessary and costly switches to second-line treatment for patients with false-positive "screening" tests (i.e., having an undetectable VL). The score was assessed in a patient population in Uganda, showing a fairly similar test performance as in the derivation study in Cambodia; however, the sample size was relatively small.[10]

Validation of the score in the setting for which it was originally intended to be used (internal validation) has not been reported yet. Moreover, validation in patients on ART for prolonged periods, by now an increasingly large population in most ART programs, is necessary as well. Furthermore, to what extent clinicians can easily and correctly apply this score in routine care remains to be assessed.

We now conducted a study in Cambodia to internally validate the algorithm, to explore its cost-saving potential, and to assess its operational performance when used in a routine setting.

METHODS

Study Setting

SHCH is a nongovernmental hospital in Phnom Penh, Cambodia. Since 2003, the hospital provides ART at no cost as part of the national program. Patients were initiated and treated according to WHO recommendations.[6,9] Patients were seen at regular intervals for clinical and laboratory monitoring and adherence assessment. Within routine care, VL was only performed for suspicion of treatment failure, using the CPS previously published.[7] Details of the program have been described before.[11,12]

Study Design and Inclusion Criteria/ Study Population

We conducted a cross-sectional study within an established ART cohort in Cambodia between May 10, 2010, and June 3, 2011. All consenting adult individuals on a nonnucleoside containing (standard) first-line regimen for a minimum of 1 year and with routine laboratory monitoring results available in the previous 6 months (CD4 cell counts and hemoglobin levels) were invited to participate.

Study Procedures

After obtaining informed consent, 2 visits were planned as follows: a first visit was 6 months after the previous routine laboratory testing for determination of follow-up CD4 cell counts and hemoglobin levels. In addition, a VL test was done. A second visit was planned 1 month later. At this point in time, the patient was clinically evaluated as per routine clinical care, adherence was assessed by the adherence counselor, and laboratory data were reviewed. At the end of the visit, the attending physician calculated the prediction score (individual items and total score). Subsequently, the VL result was made available to the physician.

Data Collection

The following data were extracted from the patient file by the principal investigator (V.P.), blinded to the VL result, and stored in a dedicated study database: sex, age, details of current and previous ART, number of occasions treatment was substituted for all reasons (excluding treatment failure) or due to the start of tuberculosis treatment, number of interruptions of treatment (no drugs at all), whether "controlled stop" (tail with bitherapy) was done when treatment interruption, new onset or recurrence of papular pruritic eruption (PPE), WHO clinical T stage (highest) over the last 6 months, body weight, CD4 cell count, hemoglobin, mean corpuscular volume (MCV), and total lymphocyte count (at time of outcome ascertainment and 6 month before); pill count and visual analogue scale for adherence assessment.

Laboratory Testing

Hemoglobin, MCV, and total lymphocyte count were measured on site by Sysmex KX-21 (Sysmex Corporation, Kobe, Japan). The CD4 cell count was performed at the National Institute of Public Health in Phnom Penh, Cambodia, by FACSCount (Becton Dickinson, Franklin Lakes, NJ). VL testing was done at the Institute Pasteur Du Cambodge on −80°C frozen plasma using the ANRS second-generation (G2) real-time polymerase chain reaction test [G2 Generic HIV-1 viral load ANRS Kit, Biocentric, Bandol, France] This in-house method allows quantification of HIV-1 B and non–B subtypes including the A/E subtype circulating in South East Asia. Using only 0.2 mL of plasma, the threshold of the assay is 250 copies per milliliter This real-time HIV-1 RNA VL assay is periodically validated through ANRS and National Institutes of Health external quality controls.

Data Analysis

The clinical scoring system contains 8 items and relies on clinical, laboratory, and adherence information (Fig. 1). Summing the weighted predictor scores of the individual's risk factors yields the total predictor score for each patient.

In line with the derivation study, virological failure was defined as a VL above 1000 copies per milliliter at the cross-sectional survey. The diagnostic performance of the scoring system in the validation population was assessed by calculating the sensitivity, specificity, and (positive and negative) predictive values at the different cutoffs of the score together with 95% confidence intervals (CIs) calculated using Wilson score method. The overall test performance was assessed by calculating the area under the receiver-operating characteristic (AUROC) curve and 95% CIs calculated using the bootstrap method. We designed the study using a similar approach to that of clinical noninferiority trials by requiring that the scoring system does not show clinically important losses of diagnostic accuracy in the validation study, compared with the original derivation study. In the original study, the area under the curve (AUC) of the scoring system was 70%. The scoring system was considered validated in SHCH, if the 95% CI for the AUC was entirely above 60%, that is, a maximal loss of AUC compared with the original study of 10%. The performance of the WHO clinical and immunological failure criteria was assessed as well. WHO failure criteria included new or recurrent WHO stage 4 conditions, a CD4 count below baseline, a CD4 decrease of 50% from the peak CD4 count during treatment, or a CD4 count below 100 cells per microliter at a minimum of 12 months of ART. In secondary analysis, a VL cutoff of 5000 copies per milliliter was taken to define treatment failure, as recommended by WHO.

The operational accuracy of the score was assessed by comparing the score given by the treating physician with the score calculated based on the available data at a given visit and expressed as percentage correctly scored. An analysis of types of errors was done by reviewing all scores given.

To determine the additional value of other possible predictors, not included in the original score, we assessed the likelihood ratio and statistical significance of the predictors in a logistic regression model including the original virological failure score as categorical predictor. The same methodology as for the original score derivation was used.[7]

Cost Analysis

The implementation of different monitoring strategies can have budgetary implications, mainly mediated through costs of second-line treatment and VL testing. Taking the perspective of a program manager or policy maker, we estimated the difference in cost and number of patients correctly diagnosed and treated over the period of 1 year for 3 strategies. First, the algorithm with targeted VL testing at the CSP cutoff of 2; second, routine VL monitoring; and third, WHO clinical and immunological switching criteria without VL testing. A decision tree was developed using TreeAge Pro 2012 (TreeAge Software, Inc.), assuming patients would undergo 1 evaluation for treatment failure with any of the 3 strategies. Only direct costs related to VL testing and second-line treatment use over a period of 1 year were taken into consideration, based on current cost in Cambodia for VL testing ($25 per test) and the cost for second-line regimens ($400 per year). One-way sensitivity analysis was conducted on the following input variables: prevalence of treatment failure in the population (range: 0.1%–40%); cost of VL test (range: $5–$100); cost of second-line treatment (range: $50–$1000). Cost of false negatives was not taken into consideration.

Sample Size

The sample size was calculated to estimate AUC of the predictor score to within 5% (half-width of the 95% CI), assuming a prevalence of virological failure of 10%. We obtained the required sample size by simulation from the distribution of the score observed in virological failures or not-failing patients in the derivation study. The required sample size was 1200 patients, of whom, approximately 120 were expected to be viral failures. Given the lower prevalence observed in the current study, the actual precision is lower.

RESULTS

Characteristics of the Study Population

Between May 10, 2010, and June 3, 2011, a total of 1502 individuals were invited to participate in the study, of which 1500 enrolled in the study. Of these, 1490 were included in analysis (Fig. 2). The patient characteristics at baseline (ART initiation) and at the time of evaluation for treatment failure are presented in Table 1. Around half (56.6%) were female; the median age at the time of study evaluation was 38 years [interquartile range (IQR): 33–44]. The median baseline CD4 count was 94 cells per microliter (IQR: 28–205).

By the time of evaluation for treatment failure, the median time on ART was 3.6 years (IQR: 2.1–5.1) (Table 1). Median body weight was 53 kg (IQR: 48–60), the median CD4 cell count had increased to 379 cells per microliter (IQR: 265–507). A total of 45 patients (3.0%) had virological failure (VL $>$ 1000 copies/mL), including 33 individuals with a VL $>$5000 copies per milliliter. The median VL of the 45 failing individuals was 32,400 copies per milliliter (IQR: 4799–112,062). In addition, 21 individuals had a detectable VL (VL $>$ 250 copies/mL) but below the study threshold of 1000 copies per milliliter.

Validation of the Scoring System

The total score calculated for each individual patient ranged from 0 to 6. Overall, the scoring system performed well in the validation population, with an AUROC of 0.75 (95% CI: 0.67 to 0.83). This compares favorably with the AUROC of 0.70 (95% CI: 0.64 to 0.76) in the derivation population (Fig. 3). Following the decision rules in the protocol, we can consider the scoring system to be validated in the Cambodian setting (95% CI lies entirely above 60%).

As seen in Table 2, sensitivities and specificities at different cutoffs of the score were fairly comparable with those documented in the derivation population. Sensitivities tended to be higher in the validation population, specificities slightly lower. With a prevalence (pretest probability) of

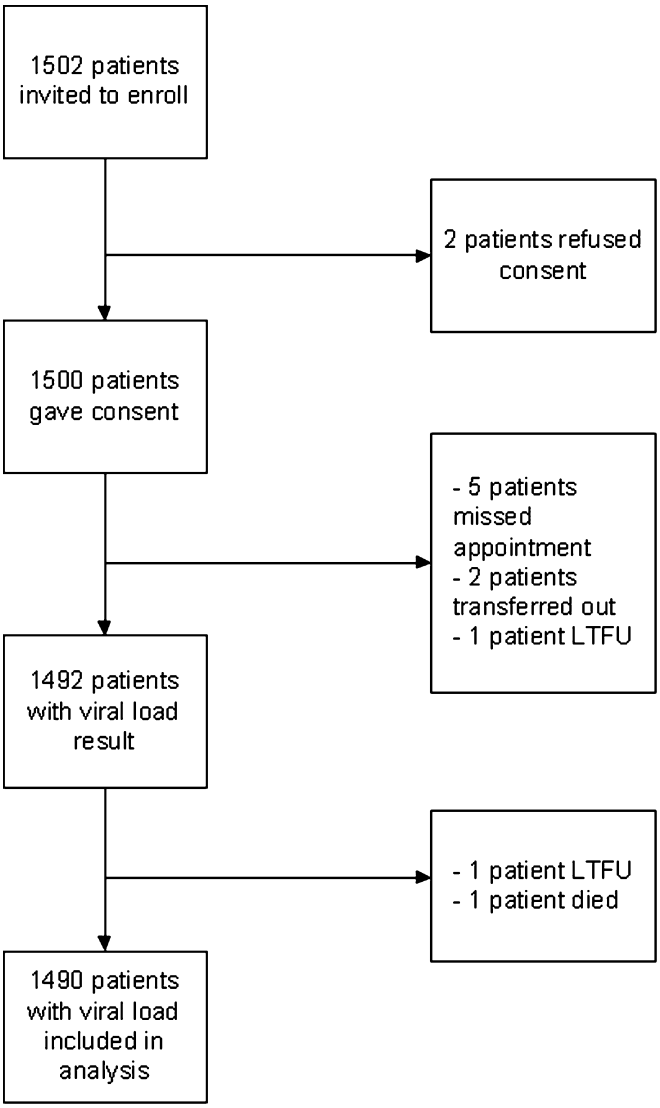

FIGURE 2. Flow chart of patients enrolled in the study. LTFU, lost to follow-up.

TABLE 1. Patient Characteristics (N = 1490)

Patient characteristics at ART initiation	
Sex: male, n (%)	646 (43.4)
Prior ART experience, n (%)	77 (5.2)
On TB treatment, n (%)	398 (26.7)
WHO clinical stage 3/4, n (%)	1151 (77.2)
CD4 (cells/μL), median (IQR)	94 (28–205)
Regimen, n (%)	
D4T-3TC-NVP	1056 (70.9)
D4T-3TC-EFV	343 (23.0)
AZT-3TC-NVP	62 (4.2)
AZT-3TC-EFV	24 (1.6)
TDF-3TC-NVP	1 (0.1)
TDF-3TC-EFV	4 (0.3)
Patient characteristics at time of evaluation for treatment failure	
Age (yrs), median (IQR)	38 (33–44)
Time on ART (years): median (IQR)	3.6 (2.1–5.1)
WHO T stage 3/4, n (%)	49 (3.3)
PPE: n (%)	23 (1.5)
VAS < 95%, n (%)—(n = 1486)	8 (0.5)
Weight (kg), median (IQR)	53 (48–60)
CD4 (cells/μL), median (IQR)	379 (265–507)
Hemoglobin (g/dL): median (IQR)	12.7 (11.5–13.9)
Regimen, n (%)	
D4T-3TC-NVP	415 (27.9)
D4T-3TC-EFV	125 (8.4)
AZT-3TC-NVP	570 (38.3)
AZT-3TC-EFV	257 (17.3)
TDF-3TC-NVP	57 (3.8)
TDF-3TC-EFV	38 (2.6)
Other	28 (1.9)

TB, tuberculosis; VAS, visual analogue scale; D4T, stavudine; 3TC, lamivudine; NVP, nevirapine: EFV, efavirenz; AZT, zidovudine; TDF, tenofovir.

treatment failure of 3%, positive predictive values of the scoring system varied from 5.8% (score ≥ 1), more than 12.8% (score ≥ 2), to 27.3% (score ≥ 4). Negative predictive values ranged from 98.9 (score ≥ 1) to 97.2% (score ≥ 4). Sensitivity and specificity of the WHO criteria was 33.3% and 92.4%, respectively. Additional predictors resulted in limited additional predictive value of the score with AUROC of the 4 alternative scores ranging between 0.77 and 0.79 ($P > 0.05$, for all comparisons with original score). All predictors included in the original score also were shown to be predictive of failure in the validation dataset.

When using the WHO-recommended VL cutoff of 5000 copies per milliliter to define treatment failure, the AUROC increased to 0.83 (95% CI: 0.77 to 0.90). At the CPS cutoff of 1, this corresponded with a sensitivity of 90.9% and specificity of 60.4%. At the CPS cutoff of 2, sensitivity and specificity was 57.6% and 90.0%, respectively. For the WHO criteria, a sensitivity of 45.4% and specificity of 92.4% was found at this VL threshold.

Operational Performance of the Score: Accuracy of the Physician-Calculated Score

The overall diagnostic accuracy of the score was reduced when using the score as calculated by the physician (AUC 0.69; 95% CI: 0.60 to 0.77) (Fig. 3). Discordance between the physician score and the study score was seen in 290 (19.5%) cases. The difference was small in most cases, with a 1-point difference in 250 cases. The physician score tended to be lower, with a median difference of 1 point (median of −1; range: −3 to +4). Differences in scoring of the individual items were most commonly seen for those items that needed more calculation and 2 or more data points like the percentage decrease from peak in CD4 count (150 discrepancies) or the decline in hemoglobin values (86 discrepancies). Discrepancies in terms of ART experience (requiring review of patient history) were also common (59 cases in total), including 52 ART-experienced individuals (as defined through complete file review) defined as not ART experienced by the physician.

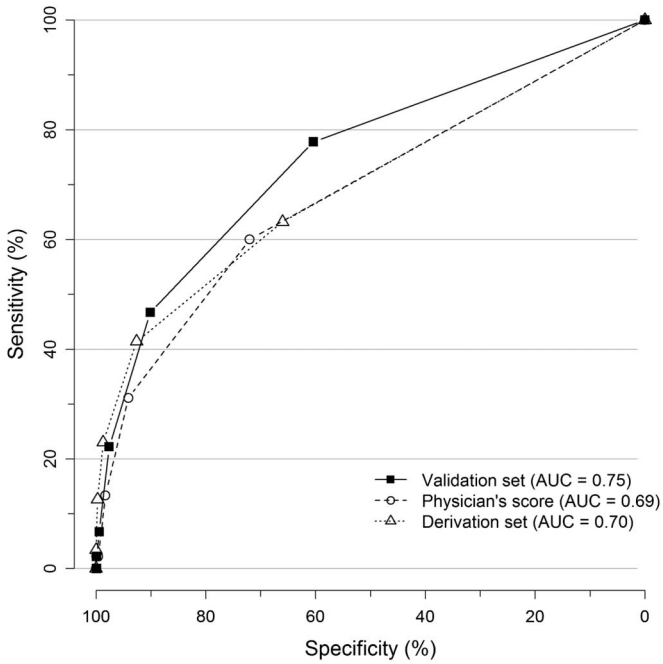

FIGURE 3. ROC curve summarizing the performance of the scoring system during development, validation, and when applied in an operational setting.

Cost of the Different Monitoring Strategies

The strategies with routine VL testing or using WHO clinical and immunological failure criteria were the most expensive. Both strategies were approximately 4 times more costly compared with a strategy with targeted VL testing based on a CPS ≥2. Although routine VL testing was the most expensive strategy, it was also the best performing in terms of diagnostic accuracy, followed by the targeted VL strategy. The strategy using the WHO criteria was "dominated" by the targeted VL strategy; that is, it was less accurate and more expensive. If the routine VL strategy were to be implemented rather than targeted VL, the incremental cost per case correctly diagnosed and treated would be US $1790 (Table 3). The findings remained the same across a plausible range of costs and prevalence of virological failure.

DISCUSSION

Improved and cost-effective treatment monitoring strategies are urgently needed, even more so given the current global budget shortfalls for HIV care and treatment.[2] Such strategies should be evidence based and carefully assessed before widespread implementation. Our algorithm combining a CPS with targeted VL performed well in validation and has cost-saving potential. Some issues with the operational performance were detected mainly in the predictors that required comparison with previous data.

With the use of our algorithm at the CPS cutoff of 2, VL tests would have been done in 11% of the patients, 46.7% of treatment failures would have been picked up. This corresponds to a number of 7.8 VL tests to detect 1 case of treatment failure. Without inclusion of targeted VL, 164 (11%) individuals would have been switched to second-line treatment, including 143 (87%) with undetectable VL. We note that the positive predictive values remained low, which partly relates to the low prevalence of virological failure in this study. Although we anticipated a prevalence of 10% in this population on ART for several years, the actual prevalence of 3% was extremely low, which is lower than most published data from Cambodia[13,14] and other low-income countries.[15] With failure rates of 10%, predictive values would have increased from 5% to 17% (score ≥ 1) and from 26% to 55% (score ≥ 4).

The CPS seemed to lose some of its performance when applied by the physicians. Based on the evaluation of discrepancies with the study score, issues with the documentation and the interpretation of trends in laboratory test results seemed to be the main underlying reason. Although not a problem in this study, it is clear that the CPS can only be applied to patients who have all the data available, which is a weakness. Especially, items like CD4 count decrease from peak values critically depend on availability and revision of all sample results while calculating the scores.

TABLE 2. Diagnostic Accuracy of Cambodian CPS and WHO Criteria to Identify Virological Failure (Validation and Derivation Population)

	Validation Population					Derivation Population	
Score	n (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
≥1	607 (40.7)	77.8 (62.9 to 88.8)	60.4 (57.8 to 63.0)	5.8 (4.1 to 7.9)	98.9 (97.9 to 99.5)	63.2 (51.4 to 73.7)	66.0 (63.2 to 68.6)
≥2	164 (11.0)	46.7 (31.7 to 62.1)	90.1 (88.5 to 91.6)	12.8 (8.1 to 18.9)	98.2 (97.3 to 98.8)	41.4 (31.2 to 52.4)	92.6 (91.2 to 93.8)
≥ 3	45 (3.0)	22.2 (11.2 to 37.1)	97.6 (96.7 to 98.3)	22.2 (11.2 to 37.1)	97.6 (96.7 to 98.3)	23.0 (15.4 to 32.9)	98.7 (98.1 to 99.2)
≥4*	11 (0.7)	6.7 (1.4 to 18.3)	99.4 (98.9 to 99.8)	27.3 (6.0 to 61.0)	97.2 (96.1 to 98.0)	12.6 (6.8 to 22.3)	99.7 (99.3 to 99.9)
WHO criteria†	125 (8.4)	33.3 (20.0 to 49.0)	92.4 (90.9 to 93.7)	12 (6.9 to 19.0)	97.8 (96.9 to 98.5)	28.7 (19.9 to 39.5)	89.7 (87.9 to 91.4)

*Only 2 patients had scores ≥5, consequently categories ≥5 and ≥6 are not shown separately.

†Defined as new or recurrent WHO stage 4 conditions, a CD4 count below baseline, a CD4 decrease of 50% from the peak CD4 count during treatment, or a CD4 count below 100 cells per microliter at a minimum of 12 months of ART.

PPV, positive predictive value; NVP, negative predictive value.

TABLE 3. Cost of Different Treatment Monitoring Strategies (N = 1490)*

	Targeted VL (CPS ≥ 2)	Routine VL	WHO Criteria
Patients correctly diagnosed and treated	1466 (98.4%)	1490 (100%)	1350 (90.6%)
Cost per correctly diagnosed and treated patient (US $)	8.5	37.1	37.0
Total annual cost (US $)	12,500	55,250	50,000
Cost of second-line treatment (US $)	8400	18,000	50,000
Cost of VL tests, (US $)	4100	37,250	0

*Based on current cost in Cambodia for VL testing ($25 per test) and additional costs for second-line regimens ($400 per year).

User-friendly medical records tools and ongoing emphasis on and monitoring of documentation practices might be of value. Because the score has been used for several years (since 2009) in the hospital, as part of a targeted VL strategy, no specific training was given on the correct use of the CPS before study implementation. Ongoing training, especially for new staff, is important to avoid errors in application.[16] The errors observed highlight that further simplification might be of benefit and at the same time reminds us of the importance of operational validation besides "statistical" validation.[17] Another strategy could be to use electronic warning systems generated by the HIV cohort database when criteria of immunological failure have been reached, in analogy to the generation of lists of defaulters after patients have missed their clinic appointments.

Only a limited number of clinical scoring systems for treatment failure have been published. One expert-based system performed poorly in validation and another (binary) system including just 2 items has never been validated.[18,19] More recently, a newly developed Ugandan system seemed promising in derivation but was reported to perform poorly in validation, worse than the Cambodian score, which was assessed in parallel.[10] Egger et al[20] have shown the usefulness of risk charts that take into account CD4 trajectories over time on ART. A range of other studies have looked at identifying risk factors for treatment failure, but did not attempt the development of a prediction score.[2] A recent study from Lesotho showed the benefit of using the Cambodian predictor score in patients who were identified as treatment failure based on WHO immunological and clinical criteria.[21] This study targeted the score to this high-risk population to identify those patients who should be switched to second-line treatment without further delay (CPS $\geq$ 5) and those who needed first confirmation by VL.

Our costing data reinforce the role of targeted VL testing—recommended in the current WHO guidelines—as a rationale way of optimizing use of scarce resources, at least while pending significant advancement in the development and validation of cheap point of care VL tests. However, the costing did not take into account the cost of false negatives. The sensitivity of the 2-step algorithm (47%) remains suboptimal, which increases the risk of delayed switching. This was also demonstrated in a clinical trial in Zambia, comparing routine VL testing with targeted VL testing. However, this delay in switch did not result in increased mortality during the first 36 months of the trial.[22] Moreover, having access to routine VL does not necessarily imply a timely switch.[23] Prolonged use of a failing regimen heralds the risk of accumulation of resistance mutations and possibly the risk of transmission of resistant viruses.[24,25] It is difficult to quantify this risk and the associated cost in a 6-month period. However, as the score will be repeated every 6 months, or earlier in case of clinical indication, this should keep the additional risk to a minimum.

Could this CPS be an alternative for the WHO failure criteria? The following arguments are in favor of the CPS. First, accuracy is (modestly) improved, at comparable cost. Although the additional number of items (and the scoring process) could possibly render it more complex, the score stresses clinical focus on a number of parameters (adherence, skin manifestations) that should regularly be assessed during patient monitoring as part of routine quality care. At the same time, it reinforces careful documentation of key patient information like treatment adherence. Moreover, one could argue that, relative to WHO T stage as a group, detection of PPE might be less error prone, more consistent across different health care settings and less relying on technical investigations. The value of PPE in predicting treatment failure has been confirmed in 2 Ugandan studies.[10,26] We acknowledge that hemoglobin monitoring might be cumbersome in some settings and comes with a certain cost, exclusion of this item had only a minor effect in the CPS performance (AUC change from 0.75 to 0.72). In contrast with a binary system as the WHO failure criteria, the scoring method allows country programs to tailor the CPS cutoff (and associated sensitivity and specificity) to their specific setting, integrating issues like prevalence of treatment failure, availability of VL, and associated financial resources. If we assume a low cost, easy to use, and high throughput VL test, threshold could be lowered to a score of 1, which would give a much higher sensitivity (77.8% and 90.9% at the VL threshold of 1000 and 5000 copies/mL, respectively) than the WHO criteria. On the other hand, a number of points argue in favor of the WHO criteria. The WHO failure criteria have been extensively validated and are now well known by most health care staff caring for HIV-infected individuals. Adoption of a novel system would again require substantial training and subsequent monitoring and follow-up. Inclusion of the WHO T stage might also increase screening for opportunistic infections.

A number of issues remain to be assessed before this algorithm can be considered for widespread implementation. To better define whether the performance of the algorithm is context specific, the algorithm should additionally be evaluated across a range of different regions, health care settings, and patient populations. Feasibility and operational validity remain to be assessed in routine settings in a number of different contexts, with clinicians using the algorithm in clinical practice. Given increased task-shifting in decentralized HIV care programs, its performance when used by nurses of other health care staff would be useful to explore. More importantly, the impact of implementing this algorithm

should be further demonstrated, to assess whether implementation of the algorithm leads to a change in behavior of the clinician with positive consequences in terms of patient outcomes and costs. Finally, the algorithm relies on the availability of VL testing. With unreliable access or long turnaround time, treatment decisions might be taken without VL result potentially leading to either inappropriate use of second-line treatment of unnecessary delays. Because its overall performance remains far from satisfactory, attempts to develop improved CPS, possibly relying on simple biomarkers, should be undertaken as well. Another consideration is the timing of the targeted VL. This score, by the intrinsic characteristics of the individual predictor items, is meant to be used in patients who are at least 1 year on ART. Early virological failure is therefore not detected. This may be a problem in patients who are nonadherent or have primary drug resistance. The risk of pretreatment drug resistance will increase as the rollout of ART continues in LMIC.[27] Pretreatment drug resistance testing is not affordable in LMIC. One could consider a systematic VL at 6 months to detect early problems with adherence or possibly primary resistance, although implementing a targeted VL thereafter.[28]

There are a number of important limitations to this study. First, treatment failure was defined on a single VL measurement, in line with the derivation study. It has been found that amongst individuals with detectable VL, VL suppression occurs in a substantial number of patients after VL testing and subsequent adherence interventions.[22,29,30] However, our objective was essentially to detect viremia at the time the "index test" was conducted. With the threshold use, viral "blips" would have been ruled out. Few patients were taking tenofovir-based ART, increasingly being used as first-line treatment in resource-constrained settings. We do not see strong reasons why the performance of our clinical score would be different when taking tenofovir, although this should ideally be confirmed. The need to further assess its operational validity and potential impact has been mentioned already. Generalizability is limited by the single-center design of the study conducted in a nongovernmental hospital in an urban setting. However, the patient population of the hospital originates from both rural and urban areas. Patients are almost universally poor, in line with the average HIV-infected patient (population) in Cambodia.[31] Still, the hospital is clearly resourced better than most hospitals in the public health system, possibly leading to higher quality of care. This could affect prevalence of treatment failure, but is unlikely to alter the biological associations between virological failure and the items in the CPS.

CONCLUSIONS

Our algorithm combining a CPS with targeted VL testing performed well in validation and has cost-saving potential. Although awaiting the development of a cheap point of care VL test, targeted VL testing combined with clinical prediction has a role to play in optimizing VL testing. Further studies to assess its performance, feasibility, and impact in different settings are warranted.

ETHICAL CONSIDERATIONS

All study participants provided written consent. The study was approved by the Institutional Review Board of the Institute of Tropical Medicine in Antwerp (Belgium), by the Ethics Committee of the University Hospital of Antwerp (Belgium), and by the National Ethics Committee for Health Research in Phnom Penh (Cambodia).

ACKNOWLEDGMENTS

The authors are grateful to the doctors and patients of SHCH for their contribution to the data collection.

REFERENCES

1. World Health Organization. *Global HIV/AIDS Response: Epidemic Update and Health Sector Progress Towards Universal Access: Progress Report, November 2011*. Geneva, Switzerland: World Health Organization; 2011.
2. Lynen L, Van Griensven J, Elliott J. Monitoring for treatment failure in patients on first-line antiretroviral treatment in resource-constrained settings. *Curr Opin HIV AIDS.* 2010;5:1–5.
3. Kahn JG, Marseille E, Moore D, et al. CD4 cell count and viral load monitoring in patients undergoing antiretroviral therapy in Uganda: cost effectiveness study. *BMJ*. 2011;343:d6884.
4. Laurent C, Kouanfack C, Laborde-Balen G, et al. Monitoring of HIV viral loads, CD4 cell counts, and clinical assessments versus clinical monitoring alone for antiretroviral therapy in rural district hospitals in Cameroon (Stratall ANRS 12110/ESTHER): a randomised non-inferiority trial. *Lancet Infect Dis.* 2011;11:825–833.
5. Mermin J, Ekwaru JP, Were W, et al. Utility of routine viral load, CD4 cell count, and clinical monitoring among adults with HIV receiving antiretroviral therapy in Uganda: randomised trial. *BMJ*. 2011;343:d6792.
6. World Health Organization. *Antiretroviral Therapy for HIV Infection in Adults and Adolescents: Recommendations for a Public Health Approach, 2006 Revision*. Geneva, Switzerland: World Health Organization; 2006.
7. Lynen L, An S, Koole O, et al. An algorithm to optimize viral load testing in HIV-positive patients with suspected first-line antiretroviral therapy failure in Cambodia. *J Acquir Immune Defic Syndr.* 2009;52:40–48.
8. Rawizza HE, Chaplin B, Meloni ST, et al. Immunologic criteria are poor predictors of virologic outcome: implications for HIV treatment monitoring in resource-limited settings. *Clin Infect Dis.* 2011;53:1283–1290.
9. World Health Organization. *Antiretroviral Therapy for HIV Infection in Adults and Adolescents, Recommendations for a Public Health Approach, 2010 Revision*. Geneva, Switzerland: World Health Organization, 2010.
10. Abouyannis M, Menten J, Kiragga A, et al. Development and validation of systems for rational use of viral load testing in adults receiving first-line ART in sub-Saharan Africa. *AIDS*. 2011;25:1627–1635.
11. Thai S, Koole O, Un P, et al. Five-year experience with scaling-up access to antiretroviral treatment in an HIV care programme in Cambodia. *Trop Med Int Health.* 2009;14:1048–1058.
12. Van Griensven J, Thai S. Predictors of immune recovery and the association with late mortality while on antiretroviral treatment in Cambodia. *Trans R Soc Trop Med Hyg.* 2011;105:694–703.
13. Ferradini L, Laureillard D, Prak N, et al. Positive outcomes of HAART at 24 months in HIV-infected patients in Cambodia. *AIDS*. 2007;21:2293–2301.
14. Pujades-Rodriguez M, Schramm B, Som L, et al. Immunovirological outcomes and resistance patterns at 4 years of antiretroviral therapy use in HIV-infected patients in Cambodia. *Trop Med Int Health.* 2011;16:205–213.
15. Barth RE, van der Loeff MF, Schuurman R, et al. Virological follow-up of adult patients in antiretroviral treatment programmes in sub-Saharan Africa: a systematic review. *Lancet Infect Dis.* 2010;10:155–166.
16. Vorkas CK, Tweya H, Mzinganjira D, et al. Practices to improve identification of adult antiretroviral therapy failure at the Lighthouse Trust clinic in Lilongwe, Malawi. *Trop Med Int Health.* 2012;17:169–176.
17. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA*. 2000;284:79–84.
18. Colebunders R, Moses KR, Laurence J, et al. A new model to monitor the virological efficacy of antiretroviral treatment in resource-poor countries. *Lancet Infect Dis.* 2006;6:53–59.

19. Meya D, Spacek LA, Tibenderana H, et al. Development and evaluation of a clinical algorithm to monitor patients on antiretrovirals in resource-limited settings using adherence, clinical and CD4 cell count criteria. *J Int AIDS Soc.* 2009;12:3.
20. Egger M, Keiser O, Zhou J, et al. CD4 cell response and virologic failure: a risk chart. Abstract 634. Presented at the 19th Conference on Retroviruses and Opportunistic Infections; March 5–8, 2012; Seatlle, WA.
21. Labhardt ND, Lejone T, Setoko M, et al. A clinical prediction score in addition to WHO criteria for antiretroviral treatment failure in resource-limited settings—experience from Lesotho. *PLoS One*. 2012;7:e47937.
22. Saag M, Westfall A, Luhanga D, et al. A cluster randomized trial of routine vs discretionary viral load monitoring among adults starting ART: Zambia. Abstract 87. Presented at the 19th conference on retroviruses and opportunistic infections; March 5–8, 2012; Seatlle, WA.
23. Johnston V, Fielding K, Charalambous S, et al. Outcomes following virological failure and predictors of switching to second-line ART: a multi-site treatment program in South Africa. Abstract 644. Presented at 19th Conference on Retroviruses and Opportunistic Infections; March 5–8, 2012; Seattle, WA. 2012.
24. Hosseinipour MC, van Oosterhout JJ, Weigel R, et al. The public health approach to identify antiretroviral therapy failure: high-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy. *AIDS*. 2009;23:1127–1134.
25. Phillips AN, Pillay D, Garnett G, et al. Effect on transmission of HIV-1 resistance of timing of implementation of viral load monitoring to determine switches from first to second-line antiretroviral regimens in resource-limited settings. *AIDS*. 2011;25:843–850.
26. Castelnuovo B, Byakwaga H, Menten J, et al. Can response of a pruritic papular eruption to antiretroviral therapy be used as a clinical parameter to monitor virological outcome? *AIDS*. 2008;22:269–273.
27. Hamers RL, Wallis CL, Kityo C, et al. HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. *Lancet Infect Dis.* 2011;11:750–759.
28. Lynen L, Fransen K, Van Griensven J, et al. Pretreatment HIV-1 drug resistance testing in sub-Saharan Africa. *Lancet Infect Dis.* 2012;12:911.
29. Coetzee D, Boulle A, Hildebrand K, et al. Promoting adherence to antiretroviral therapy: the experience from a primary care setting in Khayelitsha, South Africa. *AIDS*. 2004;18(suppl 3):S27–S31.
30. Castelnuovo B, Sempa J, Agnes KN, et al. Evaluation of WHO criteria for viral failure in patients on antiretroviral treatment in resource-limited settings. *AIDS Res Treat.* 2011;2011:736938.
31. Sopheab H, Saphonn V, Chhea C, et al. Distribution of HIV in Cambodia: findings from the first national population survey. *AIDS*. 2009;23:1389–1395.

