



Published in final edited form as:

Trop Med Int Health. 2009 October ; 14(10): 1220–1225. doi:10.1111/j.1365-3156.2009.02338.x.

Accuracy of WHO CD4 cell count criteria for virological failure of antiretroviral therapy

Olivia Keiser¹, Patrick MacPhail², Andrew Boulle³, Robin Wood⁴, Mauro Schechter⁵, François Dabis⁶, Eduardo Sprinz⁷, and Matthias Egger¹ for the ART-LINC Collaboration of the International Databases to Evaluate AIDS (IeDEA)*

¹Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
²Clinical HIV Research Unit, Department of Medicine, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa
³School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa
⁴The Desmond Tutu HIV Centre, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
⁵Hospital Universitario Clementino Fraga Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
⁶INSERM U.593, ISPED, Université Victor Segalen, Bordeaux, France
⁷Hospital de Clinicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Summary

OBJECTIVES—To examine the accuracy of the World Health Organization immunological criteria for virological failure of antiretroviral treatment.

METHODS—Analysis of 10 treatment programmes in Africa and South America that monitor both CD4 cell counts and HIV-1 viral load. Adult patients with at least two CD4 counts and viral load measurements between month 6 and 18 after starting a non-nucleoside reverse transcriptase inhibitor-based regimen were included. WHO immunological criteria include CD4 counts persistently <100 cells/ μ l, a fall below the baseline CD4 count, or a fall of >50% from the peak value. Virological failure was defined as two measurements $\geq 10\,000$ copies/ml (higher threshold) or ≥ 500 copies/ml (lower threshold). Measures of accuracy with exact binomial 95% confidence intervals (CI) were calculated.

RESULTS—A total of 2009 patients were included. During 1856 person-years of follow up 63 patients met the immunological criteria and 35 patients (higher threshold) and 95 patients (lower threshold) met the virological criteria. Sensitivity [95% confidence interval (CI)] was 17.1% (6.6–33.6%) for the higher and 12.6% (6.7–21.0%) for the lower threshold. Corresponding results for specificity were 97.1% (96.3–97.8%) and 97.3% (96.5–98.0%), for positive predictive value 9.5% (3.6–19.6%) and 19.0% (10.2–30.9%) and for negative predictive value 98.5% (97.9–99.0%) and 95.7% (94.7–96.6%).

CONCLUSIONS—The positive predictive value of the WHO immunological criteria for virological failure of antiretroviral treatment in resource-limited settings is poor, but the negative predictive value is high. Immunological criteria are more appropriate for ruling out than for ruling in virological failure in resource-limited settings.

Keywords

highly active antiretroviral therapy; treatment failure; CD4 lymphocyte count; viral load; diagnostic techniques and procedures; Africa

Introduction

In industrialized countries, and increasingly in low-income countries, the prognosis of HIV infection has improved substantially with the introduction of potent antiretroviral combination therapy (ART) (Egger *et al.* 1997; Braitstein *et al.* 2006; Keiser *et al.* 2008a). However, with increasing exposure to ART the risk of viral resistance and subsequent treatment failure has become more important, and switching to second-line regimens is increasingly needed (Keiser *et al.* 2008a, in press; Pujades-Rodriguez *et al.* 2008).

In high-income countries the diagnosis of treatment failure and the decision to switch therapy is largely based on plasma viral load monitoring and resistance testing (Hammer *et al.* 2008). In resource-limited settings, most ART programmes do not have access to viral load testing, but rely on CD4 cell counts and clinical criteria. The World Health Organization (WHO) therefore developed immunological and clinical criteria for treatment failure to guide decisions on when to switch to second-line regimens (World Health Organization 2006). We analysed data from ART programmes in resource-limited settings that monitor both CD4 cell counts and viral load to examine sensitivity, specificity and positive and negative predictive values of the WHO immunological criteria for virological failure of ART.

Methods

The ART-LINC collaboration of IeDEA

The ART in Lower Income Countries collaboration of the International epidemiological Databases to Evaluate AIDS (ART-LINC of IeDEA) is a collaborative network of 17 ART programmes in Africa, Latin America and Asia, which has been described in detail elsewhere (Dabis *et al.* 2005; Keiser *et al.* 2008b). Briefly, programmes from resource-constrained settings that systematically collect data on patient characteristics and treatment outcomes were eligible for participation in ART-LINC. For the present study, we included all 10 programmes that routinely monitor viral load as well as CD4 counts. Routine viral load monitoring was defined as at least one viral load measurement between 3 and 9 months after starting ART in at least 50% of patients treated at that site. The sites were located in Senegal (Dakar), Uganda (Kampala), South Africa (Cape Town: Gugulethu and Khayelitsha; Johannesburg and Soweto), Morocco (Casablanca), Argentina (Buenos Aires) and Brazil (Rio de Janeiro and Porto Alegre). In all sites Institutional Review Boards approved participation in ART-LINC.

Inclusion criteria and definitions

Since WHO recommends switching to a second-line regimen only after at least 6 months of first-line ART (World Health Organization 2006) we included all ART-naïve patients with two or more CD4 cell counts and viral load measurements between month 6 and 18 after starting ART, who were aged 16 years and older and started ART with a non-nucleoside reverse transcriptase inhibitor (NNRTI)-based regimen. For the purposes of this study, the WHO immunological criteria for treatment failure used were a decline in the CD4 cell count to the baseline value or below, a decline of at least 50% from the highest count on treatment or a persistent CD4 cell count below 100 cells/ μ l after 6 months of ART (World Health

Organization 2006). Virological failure was defined as a viral load of $\geq 10\,000$ copies/ml (higher threshold) or as a viral load of ≥ 500 copies/ml (lower threshold).

Statistical analysis

We calculated sensitivity, specificity and positive and negative predictive values with binomial exact confidence intervals for the higher and lower viral load thresholds. The first two measurements in the period between month 6 and 18 after starting ART were considered. In a first analysis, we required both measurements to meet the immunological and virological criteria: in practice many patients switch therapy only after failure has been confirmed by a second CD4 cell count or viral load measurement. The date of the second measurement was taken as the date of meeting criteria. In a further analysis only one value meeting the criteria was required. All analyses were performed in STATA version 10.1 (Stata Corporation, College Station, TX, USA).

Results

Figure 1 shows that of 11 044 treatment naïve patients aged 16 years or older, 2009 (18.2%) patients met the inclusion criteria. About the same number of patients were excluded because of an insufficient number of viral load measurements among those who did not meet immunological criteria for failure, compared to those who met criteria for immunological failure: 397 of 2343 patients (16.9%) compared to 12 of 75 patients (16.0%).

The number of patients analysed at each site ranged from 37 to 660. Table 1 shows the patient characteristics at the start of ART. The majority of patients were women from Africa. The median age was 34 years and the median year of starting ART was 2004. Treatment was started at a median CD4 cell count of 101 cells/ μ l and a median viral load of 5.0 log copies/ml. During a total of 1856 person-years of follow up from month 6 to 18 after starting ART, 4759 CD4 counts and 4618 viral load measurements were recorded. Sixty-three patients met the WHO immunological criteria and 35 patients (higher threshold) and 95 patients (lower threshold) met the virological criteria for failure.

Table 2, which summarizes the accuracy of the WHO immunological criteria for virological failure, shows that sensitivity was low, ranging from 12.6% to 48.1% depending on the definition chosen for virological failure (higher or lower threshold) and whether two or only one measurement were required to meet the criteria for failure. Specificity was higher, ranging from 86.8% to 97.3%. The positive predictive value was very poor (9.5–28.7%), whereas negative predictive values were high (88.4–98.5%).

Discussion

Viral load monitoring is the gold standard used in high-income countries to diagnose failure of ART, but it is not generally available in resource-limited settings. CD4 cell counts and clinical outcomes are used to monitor treatment in the absence of viral load. Our results show that the positive predictive value of the immunological (CD4 cell count) criteria for failure defined by WHO is poor for virological failure: depending on the definitions chosen, only about 10–30% of patients who met immunological criteria had virological failure. The negative predictive value was, however, considerably higher. For example, 95.7% of patients with CD4 counts not meeting the immunological criteria for failure had viral loads <500 copies/ml.

Our study has a number of strengths. Although most patients treated at sites participating in the ART-LINC of IeDEA network do not have access to routine viral load monitoring, a minority of patients did, and they made this study possible. More than 2000 patients from

six countries in Africa and South America could be included in the analysis. A limitation of our study is that we could not consider clinical outcomes. Some sites do not systematically collect data on clinical events and in sites that record clinical events the diagnostic capacities and definitions vary. It was therefore not possible to examine the correlation between clinical criteria and the laboratory criteria considered in our study. Notably, clinical failures were rare in a treatment programme in three African countries that included monitoring of CD4 cell counts and viral load (Palombi *et al.* 2009). It is therefore unlikely that the inclusion of clinical failure would have substantially changed our results. Finally, we stress that the sites included in our study may not be typical for all sites providing ART in these countries: they represent a sample of programmes with electronic medical record systems (Forster *et al.* 2008) and access to viral load monitoring. Their rate of immunological failure was, however, similar to that observed in ART-LINC sites without access to viral load monitoring (data not shown).

Several previous studies have shown poor concordance between CD4 and viral response. A study from the USA showed that a lack of increase in CD4 cell counts at one year had a sensitivity of 35% and a specificity of 94% in predicting viral load suppression (Moore *et al.* 2006). A study of South African gold miners found that WHO clinical and CD4 criteria had both poor sensitivity and poor specificity in detecting virological failure (Mee *et al.* 2008). Similarly, a study from Botswana found that an increase in CD4 cell count after initiating ART was only moderately accurate in identifying patients with undetectable viral load (Bisson *et al.* 2006). Recently, Badri *et al.* (2008) used data from the Cape Town AIDS cohort, where viral load was measured every 3 months, to show that CD4 count changes correlated with viral load at cohort level, but had limited utility in identifying virological failure in individual patients. An alternative method to identify treatment failure might be the assessment of adherence. A study of a private health care management programme in nine countries in Southern Africa showed that monitoring adherence may be more effective in predicting virological failure than declining CD4 cell counts (Bisson *et al.* 2008). Unfortunately, we did not have information on adherence in our study.

We focussed on the predictive values of immunological criteria, which will be most relevant to clinicians interpreting CD4 cell counts. Our study illustrates that the power of a test to rule a diagnosis in or out does not only depend on its specificity or sensitivity, as suggested by the SpPI_n (high specificity, positive, rules in) and SnNO_ut (high sensitivity, negative, rules out) rules promoted by the proponents of evidence-based medicine (Pewsnier *et al.* 2004). The power of the WHO immunological criteria to rule virological failure in was reduced dramatically by their low sensitivity, despite their high specificity. Similarly, the power to rule out depends on both sensitivity and specificity. In our study, the negative predictive values were high not because the immunological criteria were a powerful test, but because only few patients developed virological failure: predictive values will also depend on the incidence of virological failure in different programmes. Remarkably, we found in a previous study (Keiser *et al.* 2008a) that the probability of viral rebound was closely similar in South African townships and Switzerland, indicating that the rate of virological failure may not vary substantially across settings.

A limitation of our study is that it was not designed to evaluate diagnostic criteria but analysed routine clinical data, which may have introduced bias. For example, if the reference test (i.e. viral load measurement) is not applied consistently to confirm negative results of the index test (i.e. CD4 counts not indicating immunological failure), partial verification or work-up bias may be introduced (Whiting *et al.* 2004). In our study this may have led to overestimation or underestimation of predictive values (Pewsnier *et al.* 2004). However, substantial work-up bias is unlikely: the proportion of patients excluded because of an insufficient number of viral load measurements was similar among those who did not

meet immunological criteria for failure and those who did. Missing viral load data may nevertheless have affected our estimates of predictive values, but conclusions would not have changed: assuming that all patients with missing viral load measurements had virological failure would increase the positive predictive value to a maximum of 32.0% and reduce the negative predictive value to 79.5%, based on the lower threshold for virological failure and two measurements meeting criteria.

There is debate on the feasibility and cost-effectiveness of viral load monitoring in the context of scaling up of ART in resource-limited settings (Calmy *et al.* 2007; Phillips *et al.* 2008; Walensky *et al.* 2008). WHO stipulates that viral load monitoring is desirable, but not essential, for a public health approach to ART (Gilks *et al.* 2006). We recently analysed rates of switching from nonnucleoside reverse transcriptase inhibitor-based first-line regimens to protease inhibitor-based regimens in Africa, South America and Asia (Keiser *et al.* in press). We found that patients tended to switch earlier and at higher CD4 cell counts in programmes with, compared to programmes without, access to viral load monitoring. Clearly, further work is required to investigate whether other variables exist that could improve prediction of treatment failure in programmes without access to viral load monitoring, whether simplified techniques to measure viral load can be implemented, and in what intervals patients should optimally be monitored. Finally, future studies should examine the long-term clinical progression and mortality of patients meeting and not meeting the clinical, immunological and virological WHO criteria for treatment failure, and of patients switching and not switching to second-line regimens.

Acknowledgments

The ART-LINC collaboration of the International epidemiological Databases to Evaluate AIDS (IeDEA) is funded by the US National Institutes of Health (Office of AIDS Research and National Institute of Allergy and Infectious Diseases) and the French Agence Nationale de Recherches sur le Sida et les Hépatites Virales (ANRS). We are grateful to Hannock Tweya, Paula Braitstein, Martin Brinkhof, Suely Tuboi, Mar Pujades-Rodriguez, Alexandra Calmy, Nagalingeswaran Kumarasamy, Denis Nash, Andreas Jahn, Ruedi Lüthy and Mina Hosseinipour for helpful comments.

References

- Badri M, Lawn SD, Wood R. Utility of CD4 cell counts for early prediction of virological failure during antiretroviral therapy in a resource-limited setting. *BMC Infectious Diseases*. 2008; 8:89. [PubMed: 18601727]
- Bisson GP, Gross R, Strom JB, et al. Diagnostic accuracy of CD4 cell count increase for virologic response after initiating highly active antiretroviral therapy. *AIDS*. 2006; 20:1613–1619. [PubMed: 16868442]
- Bisson GP, Gross R, Bellamy S, et al. Pharmacy refill adherence compared with CD4 count changes for monitoring HIV-infected adults on antiretroviral therapy. *PLoS Medicine*. 2008; 5:e109. [PubMed: 18494555]
- Braitstein P, Brinkhof MW, Dabis F, et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet*. 2006; 367:817–824. [PubMed: 16530575]
- Calmy A, Ford N, Hirschel B, et al. HIV viral load monitoring in resource-limited regions: optional or necessary? *Clinical Infectious Diseases*. 2007; 44:128–134. [PubMed: 17143828]
- Dabis F, Balestre E, Braitstein P, et al. Antiretroviral Therapy in Lower Income Countries (ART-LINC): International collaboration of treatment cohorts. *International Journal of Epidemiology*. 2005; 34:979–986. [PubMed: 16157617]
- Egger M, Hirschel B, Francioli P, et al. Impact of new antiretroviral combination therapies in HIV infected patients in Switzerland: prospective multicentre study. *BMJ*. 1997; 315:1194–1199. [PubMed: 9393221]

- Forster M, Bailey C, Brinkhof MW, et al. Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings. *Bulletin of the World Health Organisation*. 2008; 86:939–947.
- Gilks CF, Crowley S, Ekpini R, et al. The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings. *Lancet*. 2006; 368:505–510. [PubMed: 16890837]
- Hammer SM, Eron JJ Jr, Reiss P, et al. Antiretroviral treatment of adult HIV infection: 2008 recommendations of the International AIDS Society-USA panel. *JAMA*. 2008; 300:555–570. [PubMed: 18677028]
- Keiser O, Orrell C, Egger M, et al. Public-health and individual approaches to antiretroviral therapy: township South Africa and Switzerland compared. *PLoS Medicine*. 2008a; 5:e148. [PubMed: 18613745]
- Keiser O, Anastos K, Schechter M, et al. Antiretroviral therapy in resource-limited settings 1996 to 2006: patient characteristics, treatment regimens and monitoring in sub-Saharan Africa, Asia and Latin America. *Tropical Medicine International Health*. 2008b; 13:870–879. [PubMed: 18373510]
- Keiser O, Tweya H, Boulle A, et al. Switching to second-line ART in resource-limited settings: Comparison of programmes with and without viral load monitoring. *AIDS*. 2009 Jan 15. (Epub ahead of print).
- Mee P, Fielding KL, 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. [PubMed: 18784460]
- Moore DM, Mermin J, Awor A, et al. Performance of immunologic responses in predicting viral load suppression: implications for monitoring patients in resource-limited settings. *Journal of Acquired Immune Deficiency Syndrome*. 2006; 43:436–439.
- Palombi L, Marazzi MC, Guidotti G, et al. Incidence and predictors of death, retention, and switch to second-line regimens in antiretroviral-treated patients in Sub-Saharan African sites with comprehensive monitoring availability. *Clinical Infectious Diseases*. 2009; 48:115–122. [PubMed: 20380075]
- Pewsnar D, Battaglia M, Minder C, et al. Ruling a diagnosis in or out with “SpIn” and “SnNOt”: a note of caution. *BMJ*. 2004; 329:209–213. [PubMed: 15271832]
- Phillips AN, Pillay D, Miners AH, et al. 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; 371:1443–1451. [PubMed: 18440426]
- Pujades-Rodriguez M, O’Brien D, Humblet P, Calmy A. Second-line antiretroviral therapy in resource-limited settings: the experience of Medecins Sans Frontieres. *AIDS*. 2008; 22:1305–1312. [PubMed: 18580610]
- Walensky RP, Freedberg KA, Weinstein MC. Monitoring of antiretroviral therapy in low-resource settings. *Lancet*. 2008; 372:288. [PubMed: 18657699]
- Whiting P, Rutjes AW, Reitsma JB, et al. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Annals of Internal Medicine*. 2004; 140:189–202. [PubMed: 14757617]
- World Health Organization. Antiretroviral Therapy for HIV Infection in Adults and Adolescents in Resource-limited Settings: Towards Universal Access. Recommendations for a Public Health Approach. WHO; Geneva: 2006.

Appendix

The ART-LINC of leDEA Central Coordinating Team

Eric Balestre, Martin Brinkhof, François Dabis (principal investigator), Matthias Egger (principal investigator), Claire Graber, Beatrice Fatzer, Olivia Keiser, Charlotte Lewden, Mar Pujades, Mauro Schechter (principal investigator).

Collaborating centres

ANRS 1290 (Dakar, Senegal); Adherence Monitoring Uganda (AMU) cohort; Gugulethu ART Programme, (Cape Town, South Africa); Khayelitsha ART Programme, (Cape Town, South Africa); Themba Lethu/WITS (Johannesburg, South Africa); Perinatal HIV Research Unit (Soweto, South Africa); Morocco Antiretroviral Treatment Cohort, Centre Hospitalier Universitaire (Casablanca, Morocco); Prospective Evaluation in the Use and Monitoring of Antiretrovirals in Argentina (PUMA), Buenos Aires, Argentina; South Brazil HIV Cohort (SOBRHIV), Hospital de Clinicas (Porto Alegre, Brazil); Rio de Janeiro HIV Cohort, Hospital Universitario Clementino Fraga Filho (Rio de Janeiro, Brazil).

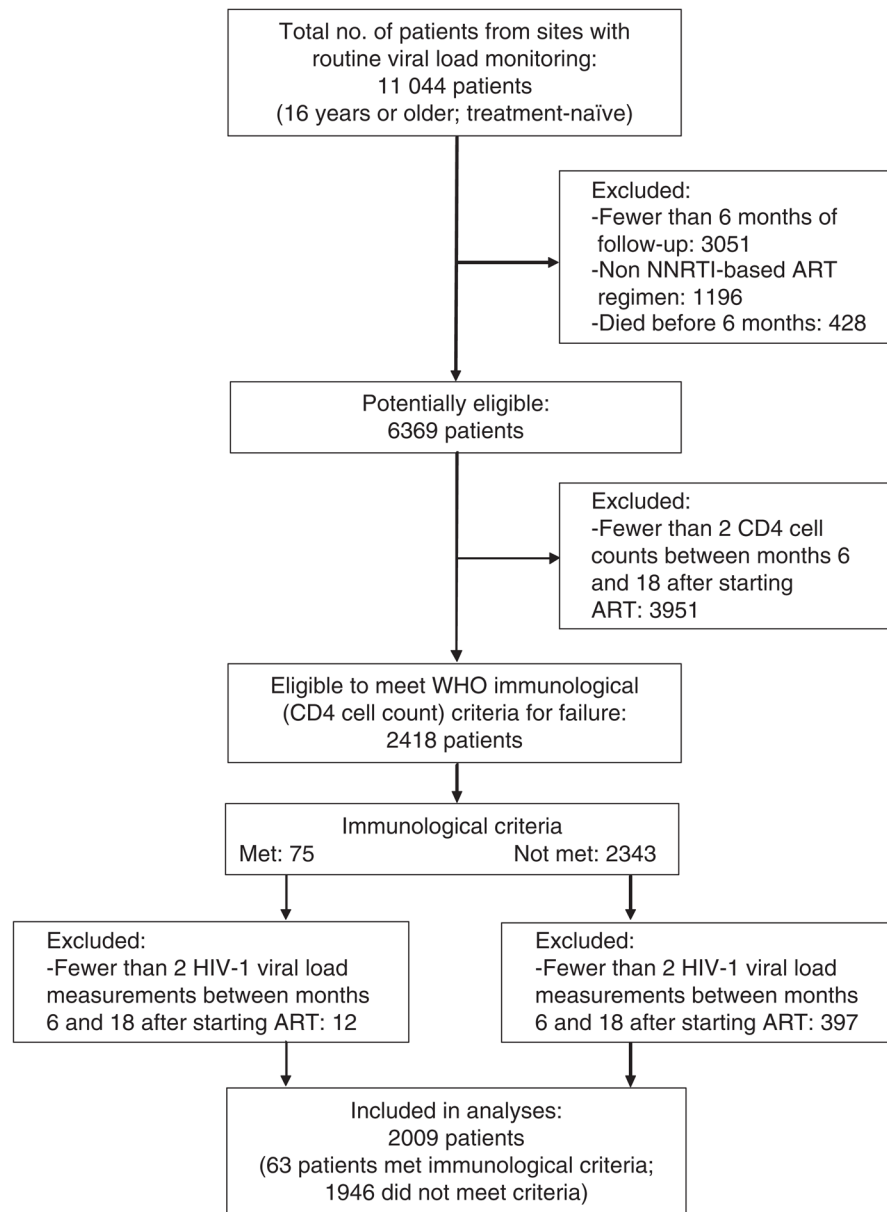


Figure 1.
Selection of study population.

Table 1

Baseline characteristics of the 2009 patients included in analyses

No. of females (%)	1283 (63.9%)
Region (%)	
Africa	1631 (81.2%)
South America	378 (18.8%)
Age (years)	
Median (IQR)	34 (30–41)
CD4 cell count (cells/ μ l)	
Median (IQR)	101 (44–168)
No. missing (%)	214 (10.7%)
HIV-1 viral load (\log_{10} copies/ml)	
Median (IQR)	5.0 (4.5–5.5)
No. missing (%)	441 (22.0%)
Clinical stage	
CDC stage C, WHO stage III/IV	1338 (71.9%)
No. missing (%)	
Year of starting ART	149 (7.4%)
Median (IQR)	2004 (2003–2004)
ART regimen	
3CT d4T EFV	925 (46.0%)
3CT AZT EFV	588 (29.3%)
3TC d4T NVP	236 (11.8%)
Other	260 (12.9%)

IQR, interquartile range; ART, antiretroviral therapy.

Table 2

Sensitivity, specificity, positive and negative predictive values of World Health Organization (WHO) immunological criteria for virological failure of antiretroviral therapy

Definition of virological failure No. of CD4 count and viral load measurements meeting criteria*	TP	FN	FP	TN	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value, % (95% CI)	Negative predictive value, % (95% CI)
10 000 copies/ml								
Based on two measurements meeting criteria	6	29	57	1917	17.1 (6.6–33.6)	97.1 (96.3–97.8)	9.5 (3.6–19.6)	98.5 (97.9–99.0)
Based on one measurement meeting criteria	76	82	245	1606	48.1 (40.1–56.2)	86.8 (85.1–88.3)	23.7 (19.1–28.7)	95.1 (94.0–96.1)
500 copies/ml								
Based on two measurements meeting criteria	12	83	51	1863	12.6 (6.7–21.0)	97.3 (96.5–98.0)	19.0 (10.2–30.9)	95.7 (94.7–96.6)
Based on one measurement meeting criteria	92	195	229	1493	32.1 (26.7–37.8)	86.7 (85.0–88.3)	28.7 (23.8–33.9)	88.4 (86.8–89.9)

Percentages with binomial exact confidence intervals are shown. TP, true positive; FN, false negative; FP, false positive; TN, true negative.

* Criteria for virological failure as shown in table, criteria for immunological failure as defined by WHO: decline in the CD4 cell count to the baseline value or below, a decline of at least 50% from the highest count on treatment or a persistent CD4 cell count below 100 cells/ μ l after six months of ART (World Health Organization 2006).