Supplementary web appendix to

Risk charts to guide targeted HIV-1 viral load monitoring of ART: development and validation in patients from resource-limited settings

Manuel Koller, PhD[a], Geoffrey Fatti, MD[b], Benjamin H Chi, MD[c], Olivia Keiser, PhD[a], Christopher J Hoffmann, MD[d], Robin Wood, MD[e], Hans Prozesky, MD[f], Kathryn Stinson, MD[g], Janet Giddy, MD[h], Portia Mutevedzi, PhD[i], Matthew Fox, PhD[j,k], Matthew Law, PhD[l], Andrew Boulle, MD, PhD[m], and Matthias Egger, MD[a,m]

on behalf of the International epidemiological Databases to Evaluate AIDS in Southern Africa (IeDEA-SA) and the TREAT Asia HIV Observational Database (TAHOD)

a) Institute of Social and Preventive Medicine, University of Bern, Switzerland
b) Kheth'Impilo, Cape Town, South Africa
c) Centre for Infectious Disease Research in Zambia, Lusaka, Zambia
d) Aurum Institute for Health Research, Johannesburg, South Africa
e) Gugulethu ART Programme and Desmond Tutu HIV Centre, University of Cape Town, South Africa
f) Division of Infectious Diseases, Department of Medicine, University of Stellenbosch and Tygerberg Academic Hospital, Cape Town, South Africa
g) Médecins Sans Frontières, Khayelitsha, Cape Town, South Africa
h) Sinikithemba Clinic, McCord Hospital, Durban, South Africa
i) Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Somkhele, South Africa
j) Health Economics and Epidemiology Research Office, University of the Witwatersrand, Johannesburg, South Africa
k) Center for Global Health & Development and Department of Epidemiology, Boston University, Boston, USA
l) Biostatistics and Databases Program, The Kirby Institute, Faculty of Medicine, The University of New South Wales, Sydney, Australia
m) Centre for Infectious Disease Epidemiology and Research, School of Public Health and Family Medicine, University of Cape Town, South Africa

Correspondence to:
Professor Matthias Egger
Institute of Social and Preventive Medicine (ISPM)
University of Bern
Finkenhubelweg 11, CH-3012 Bern, Switzerland
egger@ispm.unibe.ch

Table S1. Baseline characteristics of patients included in the development and validation of risk charts for virologic failure on antiretroviral therapy based on current CD4 count and CD4 count measured 6 months previously.

	South Africa (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)
Patients				
Number of patients	36,511 (100%)	12,909 (100%)	2,854 (100%)	1,367 (100%)
Gender				
Female	21,768 (60%)	9,142 (71%)	1,553 (54%)	427 (31%)
Male	14,743 (40%)	3,767 (29%)	1,301 (46%)	940 (69%)
Age (years)				
Median (IQR)	36 (30 – 43)	35 (30 – 42)	36 (31 – 43)	35 (30 – 43)
16 - 29	7,602 (21%)	2,882 (22%)	483 (17%)	276 (20%)
30 - 39	16,279 (45%)	5,814 (45%)	1,293 (45%)	622 (46%)
40 - 49	9,036 (25%)	3,108 (24%)	772 (27%)	332 (24%)
>= 50	3,594 (10%)	1,105 (9%)	306 (11%)	137 (10%)
CD4 count at start of ART (cells/µl)				
Median (IQR)	112 (50 – 176)	126 (71 – 173)	121 (62 – 188)	111 (37 – 203)
< 50	6,923 (25%)	1,553 (17%)	492 (21%)	382 (33%)
50 - 99	5,623 (21%)	1,902 (21%)	531 (22%)	195 (17%)
100 - 199	10,834 (40%)	4,709 (53%)	945 (39%)	322 (28%)
200 - 349	3,825 (14%)	791 (9%)	431 (18%)	266 (23%)
Year of starting ART				
Median (IQR)	2006 (2005 – 2007)	2007 (2006 – 2008)	2007 (2006 – 2007)	2004 (2002 – 2006)
Follow up time (years)				
Median (IQR)	1.92 (1.15 – 3.01)	1.99 (1.32 – 2.90)	2.97 (2.39 – 3.91)	3.14 (1.78 – 4.53)
Total	79,803	28,313	8,748	4,194
Laboratory values				
No. of triplets analysed*	135,824 (100%)	34,478 (100%)	10,041 (100%)	8,169 (100%)
No. with virologic failure	20,320 (15%)	8,269 (24%)	1,335 (13%)	792 (10%)
No. of imputed CD4 counts	13,977 (10%)	10,737 (31%)	1,877 (19%)	462 (6%)
No. of imputed VL measurements	13,644 (10%)	14,608 (42%)	5,027 (50%)	3,220 (39%)

*: CD4 count and VL measured at same time during follow-up and CD4 count measured 3 months previously.

VL, viral load

Table S2. Accuracy of prediction of virologic failure in derivation and validation cohorts. Results are shown for different cut-offs for the predicted probability of virologic failure, in the absence of targeted viral load testing.

	Model 1 (Current and baseline CD4 count)				Model 2 (Current CD4 count and CD4 count measured 6 months previously)			
	South Africa (derivation dataset)	**South Africa** (validation dataset)	**Zambia** (validation dataset)	**Asia-Pacific** (validation dataset)	**South Africa** (derivation dataset)	**South Africa** (validation dataset)	**Zambia** (validation dataset)	**Asia-Pacific** (validation dataset)
Cut-off at predicted probability > 0.2								
PPV	39%	36%	24%	20%	36%	39%	18%	16%
NPV	91%	84%	90%	93%	90%	81%	90%	92%
Sensitivity	60%	52%	71%	53%	50%	42%	56%	41%
Specificity	81%	73%	54%	74%	84%	79%	62%	78%
Cut-off at predicted probability > 0.3								
PPV	52%	43%	30%	28%	50%	46%	24%	24%
NPV	89%	82%	88%	92%	89%	79%	89%	92%
Sensitivity	43%	34%	54%	36%	30%	24%	36%	25%
Specificity	92%	87%	74%	88%	95%	91%	82%	91%
Cut-off at predicted probability > 0.4								
PPV	63%	49%	37%	39%	59%	51%	32%	34%
NPV	88%	80%	87%	91%	87%	78%	89%	91%
Sensitivity	31%	23%	42%	24%	19%	16%	25%	15%
Specificity	96%	93%	85%	95%	98%	95%	92%	97%
Cut-off at predicted probability > 0.5								
PPV	71%	56%	44%	49%	67%	56%	40%	45%
NPV	86%	80%	86%	91%	87%	78%	88%	91%
Sensitivity	23%	16%	32%	17%	13%	11%	18%	11%
Specificity	98%	96%	91%	98%	99%	97%	96%	99%
Cut-off at predicted probability > 0.6								
PPV	77%	61%	50%	58%	72%	58%	49%	52%
NPV	86%	79%	85%	90%	86%	77%	88%	91%
Sensitivity	16%	10%	23%	12%	9%	7%	11%	7%
Specificity	99%	98%	95%	99%	99%	98%	98%	99%
Cut-off at predicted probability > 0.7								
PPV	81%	65%	55%	59%	75%	55%	49%	57%
NPV	85%	78%	84%	90%	86%	76%	87%	91%
Sensitivity	10%	6%	13%	8%	5%	3%	6%	3%
Specificity	100%	99%	98%	99%	100%	99%	99%	100%
Cut-off at predicted probability > 0.8								
PPV	85%	71%	49%	65%	66%	44%	50%	40%
NPV	84%	78%	83%	90%	85%	76%	87%	90%
Sensitivity	5%	3%	4%	4%	1%	1%	1%	1%
Specificity	100%	100%	99%	100%	100%	100%	100%	100%

Table S3. Accuracy of prediction of virologic failure in derivation and validation cohorts. compared to the WHO 2010 and 2013 criteria for immunological failure, in the absence of targeted viral load testing.

	Model 1 (Current and baseline CD4 count)				Model 2 (Current CD4 count and CD4 count measured 3 months previously)			
	South Africa (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)	**South Africa** (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)
0% VL testing*								
PPV	61%	48%	35%	37%	56%	49%	28%	29%
NPV	88%	81%	87%	91%	88%	79%	89%	92%
Sensitivity	33%	24%	43%	25%	24%	19%	29%	18%
AUC	0.77	0.67	0.69	0.70	0.74	0.64	0.63	0.65
2006 WHO criteria (fall of the CD4 counts to baseline (or below) or 50% fall from on-treatment peak value or persistent CD4 levels below 100 cells/µl)								
PPV	56%	48%	32%	39%	49%	47%	25%	32%
NPV	88%	81%	86%	91%	89%	81%	90%	92%
Sensitivity	36%	25%	35%	28%	38%	34%	44%	30%
Specificity	94%	92%	84%	95%	93%	88%	79%	93%
2013 WHO criteria (fall of the CD4 counts to baseline (or below) or persistent CD4 levels below 100 cells/µl)								
PPV	56%	49%	33%	41%	49%	49%	25%	35%
NPV	87%	80%	86%	91%	88%	80%	89%	92%
Sensitivity	28%	21%	31%	24%	29%	27%	34%	27%
Specificity	96%	93%	87%	96%	95%	91%	85%	95%

* Using a probability cut-off of 0.38 for Model 1 and 0.36 for Model 2.

Figure S1. Flow chart of identifying eligible patients.

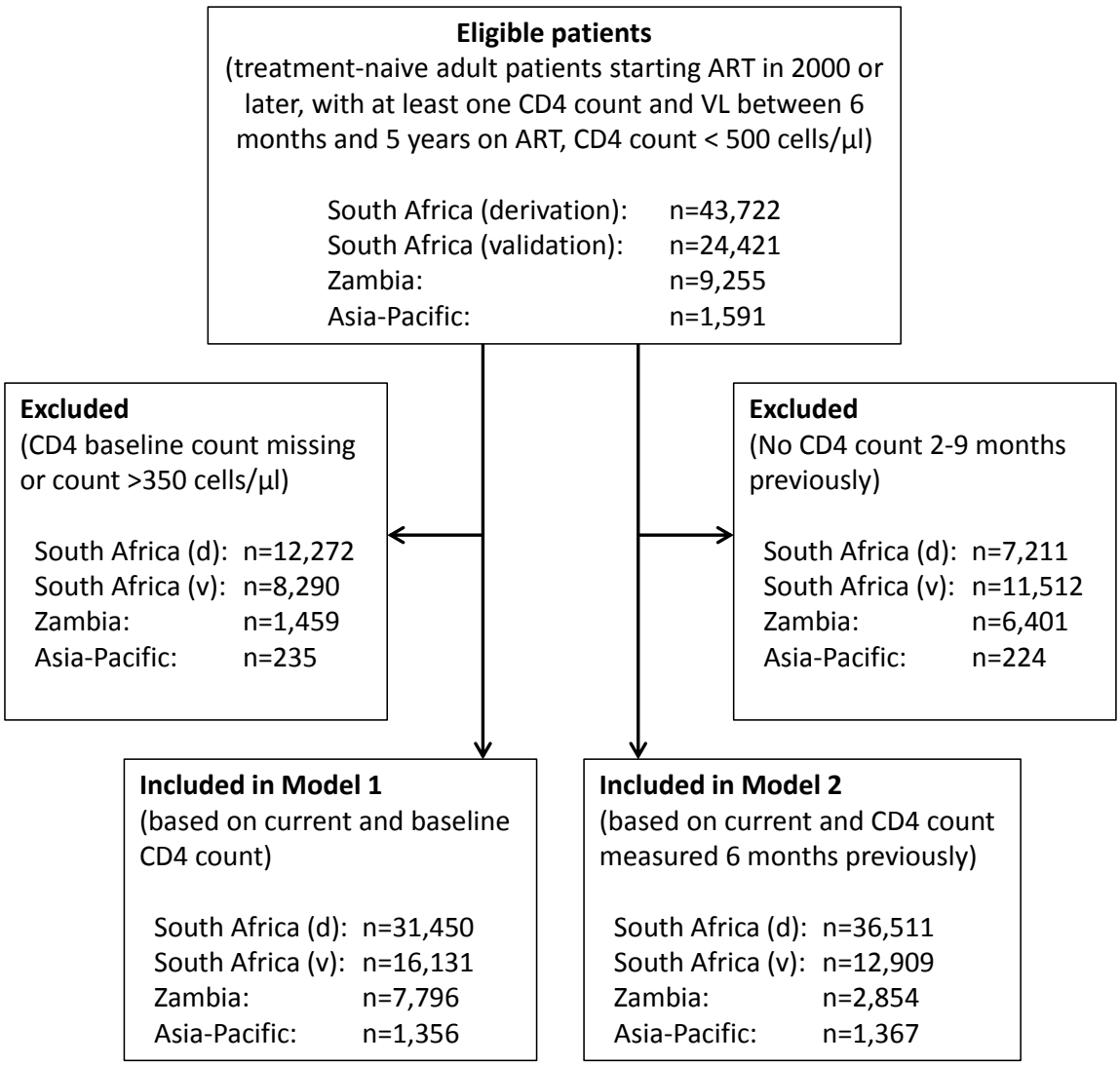


Figure S2. Risk chart for virologic failure based on baseline CD4 cell count and current CD4 cell count stratified by time on antiretroviral therapy (columns) and gender (rows).

The area between two lines of the same style contains the patients that are optimally tested given the resources available.

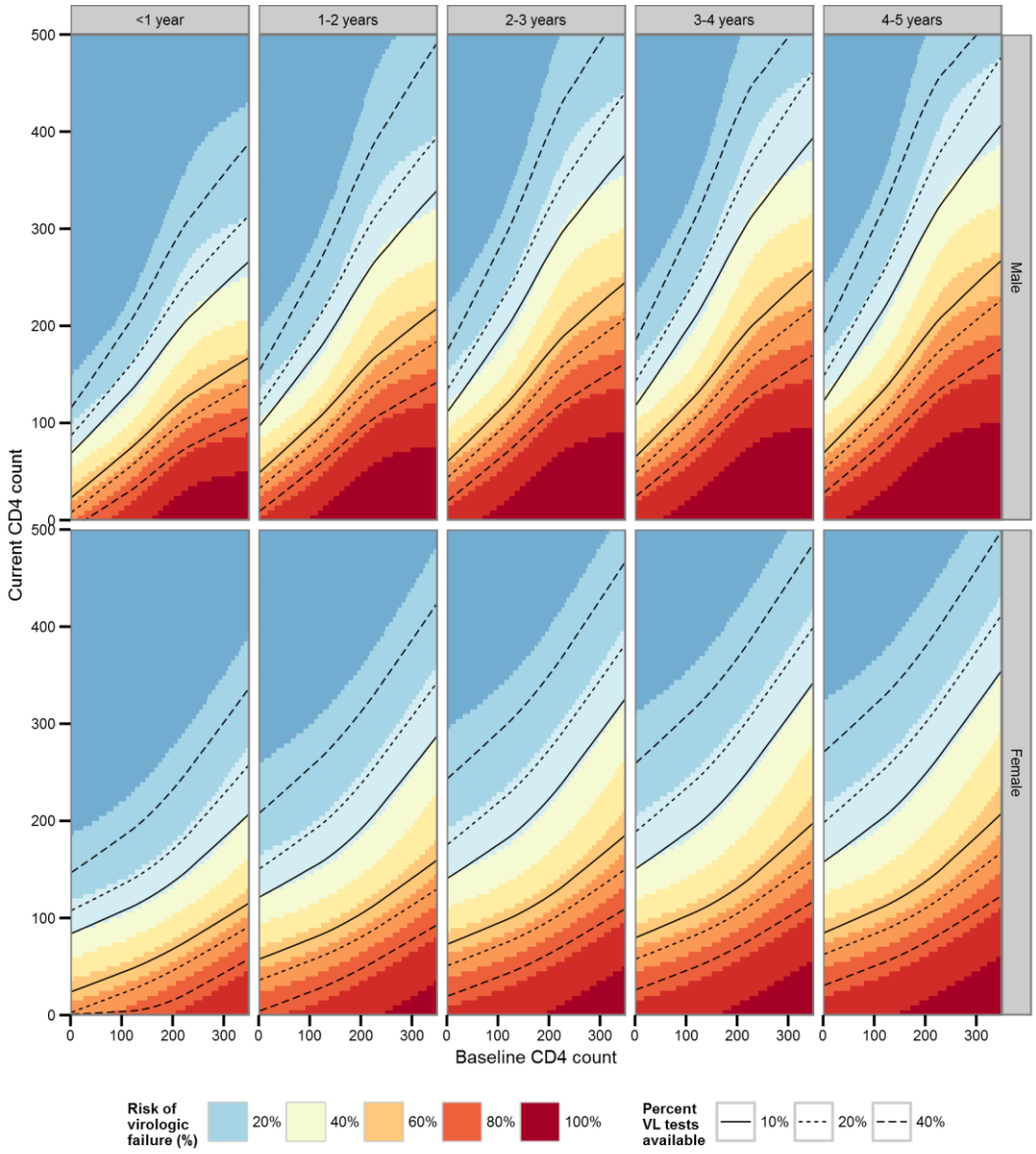
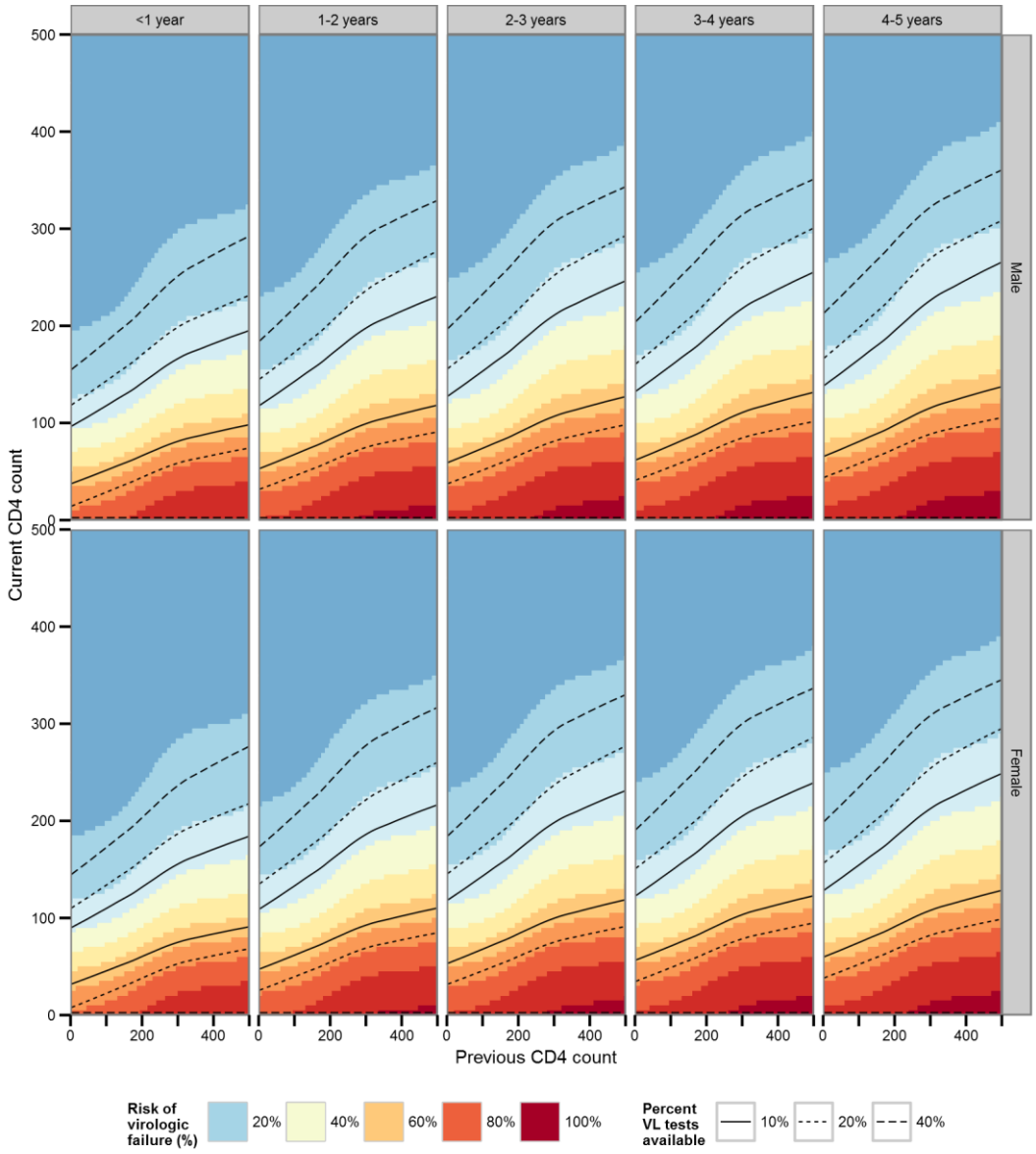


Figure S3. Risk chart for virologic failure based on CD4 cell count measured 6 months earlier and current CD4 cell count stratified by time on antiretroviral therapy (columns) and gender (rows).

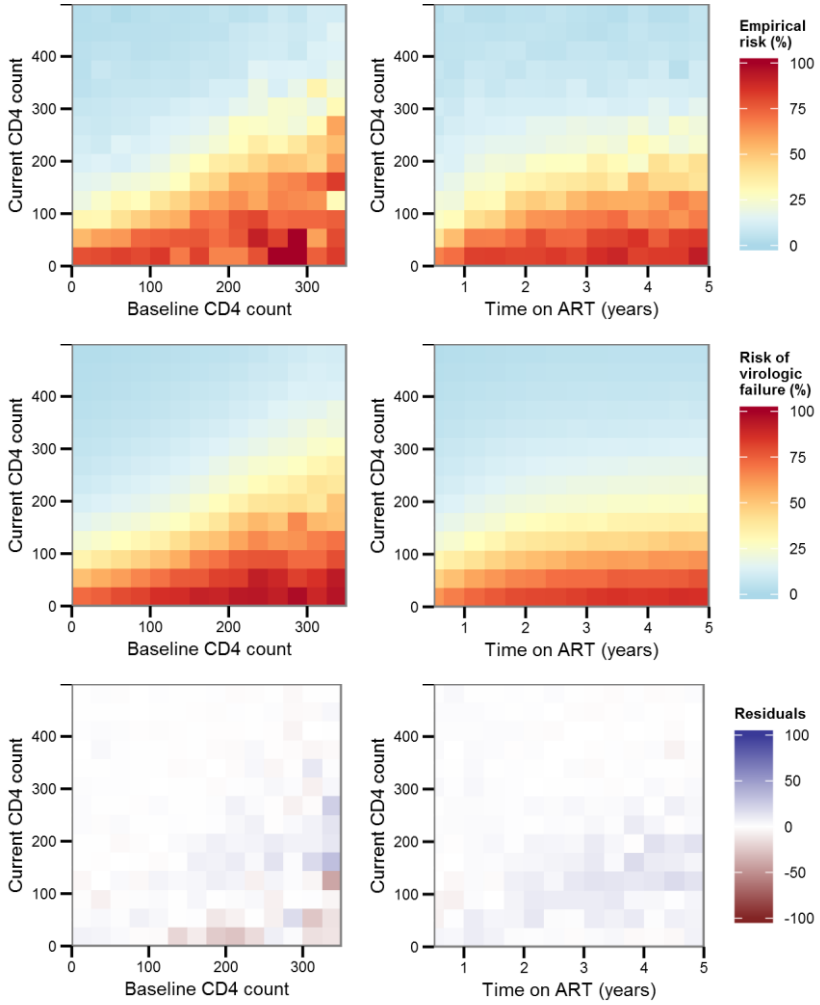
The area between two lines of the same style contains the patients that are optimally tested given the resources available.



Technical Appendix

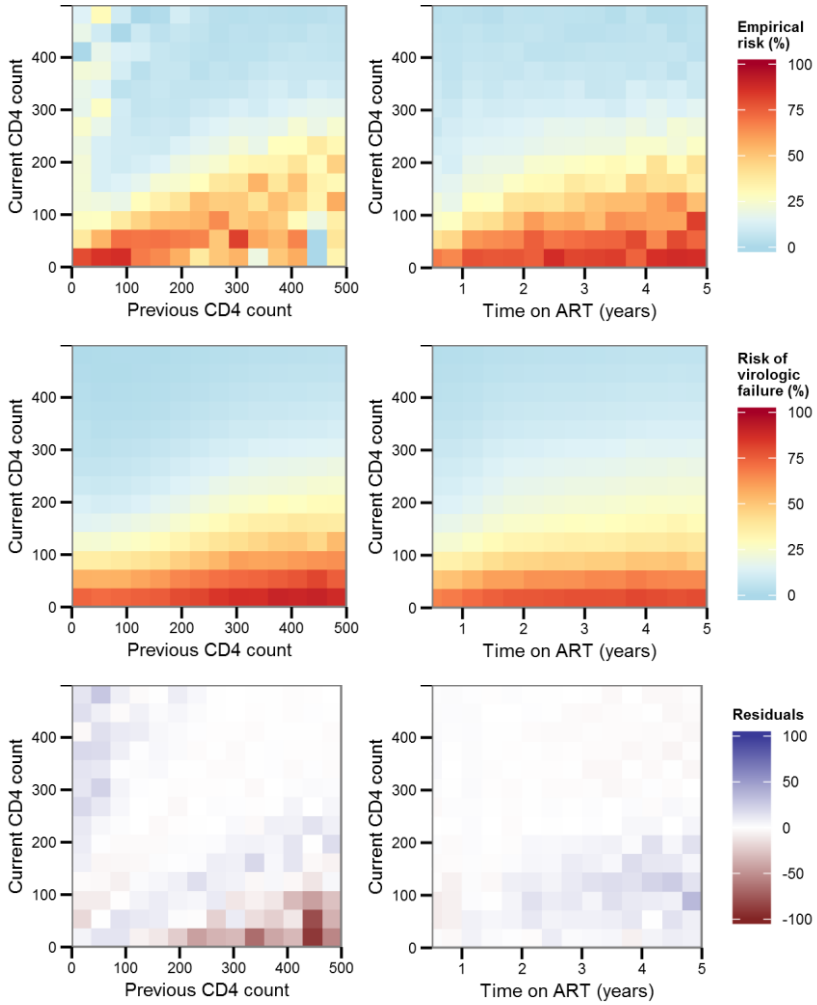
Model fit

The observed and predicted risk of virologic failure and the differences between observed and predicted risk (the residuals) in the 225 cells defined by current and baseline CD4 counts or current CD4 count and time on ART (model 1) are shown in the goodness of fit plots below. The top panels show the observed risk computed for the derivation dataset for baseline CD4 cell count against current CD4 cell count (left panels) and time on antiretroviral therapy (ART) against current CD4 cell count (right panels). The middle panels show the predicted values and the bottom panels the difference between observed and predicted risk.



Above the diagonal, the residuals are uniformly small. Immediately below the diagonal, the risk is slightly underestimated by the model and, further below, slightly overestimated.

The corresponding data for the model based on the CD4 count measured 6 months prior to the current count (model 2) are shown below. The pattern is similar but the observed risk of virologic failure was, surprisingly, increased in patients with CD4 counts below 100 cells/µl whose CD4 count increased by more than 150 cells/µl within 6 months. These were based on a small number of patients and might reflect the play of chance (observed in 888 patients out of 36,511, 2.4%).



Sensitivity analyses

We ran three sensitivity analyses. We assessed the influence of the missing values in the data in two sensitivity analyses. In a third sensitivity analysis we checked the influence of patients having multiple observations.

Missing values

For the data at hand, a value is said to be missing if there is only either a CD4 or a viral load measurement available for a certain day. The missing values are not expected to have a big impact as they are expected not to be missing at random. CD4 or viral load measurements might have been skipped as the patient was showing good health and it was deemed to be unnecessary to measure both values, especially when the patient was on successful treatment for a long time. We validated this claim by comparing the imputed data analysis with the complete case analysis (Sensitivity analysis 1).

The simple (interpolation only) imputation scheme used in this work might be problematic as it does artificially reduce the amount of variation in the data. This then usually leads to an underestimation of the variability of the estimates. For this analysis, we do not expect this to be a problem as we are not interested in p-values and the like but in predictive statistics. The predictive statistics are evaluated using data that was not used for fitting the models but had to be imputed as well. To

validate this claim, we imputed the data again adding also random errors to the imputed values (Sensitivity analysis 2).

Multiple observations per patient

In the main analysis, we treated multiple observations of one patient as independent. Doing this usually leads to an underestimation of standard errors and p-values that are much too small. The fitted estimates itself are in general not much influenced by ignoring these dependencies. We validated our methods using data that was not used in the fitting procedure and do not depend on standard errors but on predictive statistics. One possible effect of ignoring the dependence of observations of the same patient could be on the predictive statistics. For example, on one hand, patients that are on a successful treatment for a long time contribute many measurements and should be "easy" for any method to predict. On the other hand, patients that fail quickly might not yet show a decrease in CD4 and hence "hard" to predict. The same patients are more likely to have short trajectories. Such a mechanism might unduly improve the predictive statistics. We can assess whether there is a mechanism as described or similar by weighting the patients, such that all the patients have the same weight in the computation of the predictive statistics (Sensitivity analysis 3).

Methods

Sensitivity analysis 1: complete case analysis

In the complete case analysis, we did not impute any missing values. We only tried to match pairs with missing measurements that were taken less than one month apart from one another (one missing the CD4 and the other the viral load measurement).

Sensitivity analysis 2: alternative imputation method with added random errors

The artificial measurement error was added as follows. For the CD4 measurements, we took the fourth roots of the interpolated values and added random normal errors with mean 0 and standard deviation .64 and then transformed the resulting values back using the fourth power. This translates to the expectation of a true CD4 count of 750 varying between 500 and 1000 cells/µl.

Similarly for the viral load measurements, we took the base-10 logarithm, added standard normal errors and transformed back using the 10^{th} power. This translates to the expectation of a true viral load of 100,000 to be varying between 10,000 and 1,000,000 copies/ml.

Sensitivity analysis 3: multiple observations per patient

We computed weighted predictive statistics. The weights were chosen such that the weights of all observations summed to one, i.e., all the patients had the same total weight in the analysis.

Results

We show results analogous to the tables in the main paper. The results of sensitivity analysis 1 are shown in Table S4, for sensitivity analysis 2 in Table S5 and for sensitivity analysis 3 in Table S6.

The differences are generally very small ranging from -3% to +3%. The most extreme differences range up to 15% (PPV of Model 1 for Asia-Pacific data, sensitivity analysis 3) giving a much better result in the sensitivity analysis. The figures show some slight differences. All in all, the conclusions remain stable and the variation from the main analysis remains in the expected statistical error range.

Table S4. Accuracy of prediction of virologic failure in derivation and validation cohorts. Results are shown for no VL testing and for the testing of 20% or 40% of patients, using optimal rules for the range of patients tested based on the predicted probability of virologic failure.

Results from complete case analysis.

	Model 1 (Current and baseline CD4 count)				Model 2 (Current CD4 count and CD4 count measured 3 months previously)			
	South Africa (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)	**South Africa** (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)
				0% VL testing*				
	Probability cut off: 0.38				Probability cut off: 0.34			
PPV	63%	48%	45%	36%	57%	46%	24%	28%
NPV	89%	84%	89%	94%	89%	82%	93%	94%
Sensitivity	34%	27%	48%	30%	25%	22%	24%	17%
Specificity	96%	93%	88%	95%	97%	93%	93%	97%
				10% VL testing				
	Probability range tested: 0.27 - 0.53				Probability range tested: 0.24 - 0.59			
PPV	86%	75%	69%	71%	92%	85%	71%	82%
NPV	91%	87%	92%	95%	91%	85%	95%	95%
Sensitivity	47%	40%	59%	46%	38%	33%	35%	30%
Specificity	99%	97%	95%	98%	99%	98%	99%	99%
% tested	10%	14%	18%	12%	10%	14%	15%	9%
				20% VL testing				
	Probability range tested: 0.2 - 0.62				Probability range tested: 0.18 - 0.71			
PPV	94%	87%	82%	86%	97%	95%	86%	97%
NPV	93%	89%	94%	96%	93%	87%	95%	96%
Sensitivity	60%	52%	71%	59%	53%	45%	44%	42%
Specificity	99%	98%	97%	99%	100%	99%	99%	100%
% tested	20%	25%	33%	24%	20%	25%	27%	20%
				40% VL testing				
	Probability range tested: 0.12 - 0.78				Probability range tested: 0.12 - 0.96			
PPV	99%	97%	94%	95%	100%	100%	100%	100%
NPV	96%	93%	96%	98%	96%	91%	97%	97%
Sensitivity	76%	70%	82%	78%	72%	63%	66%	63%
Specificity	100%	99%	99%	100%	100%	100%	100%	100%
% tested	40%	47%	59%	47%	40%	44%	51%	39%

Table S5. Accuracy of prediction of virologic failure in derivation and validation cohorts. Results are shown for no VL testing and for the testing of 20% or 40% of patients, using optimal rules for the range of patients tested based on the predicted probability of virologic failure.

Results from analysis using imputed values, adding random normal errors to the imputed values used in the main analysis.

	Model 1 (Current and baseline CD4 count)				Model 2 (Current CD4 count and CD4 count measured 3 months previously)			
	South Africa (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)	**South Africa** (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)
0% VL testing*								
	Probability cut off: 0.38				Probability cut off: 0.37			
PPV	59%	44%	35%	40%	54%	46%	35%	35%
NPV	86%	74%	82%	85%	86%	69%	81%	86%
Sensitivity	31%	22%	39%	18%	20%	15%	21%	12%
Specificity	95%	89%	79%	94%	97%	91%	89%	96%
10% VL testing								
	Probability range tested: 0.29 - 0.5				Probability range tested: 0.27 - 0.54			
PPV	82%	70%	56%	70%	88%	80%	77%	80%
NPV	89%	77%	86%	87%	88%	73%	85%	88%
Sensitivity	42%	31%	50%	29%	32%	25%	35%	23%
Specificity	98%	94%	89%	97%	99%	97%	97%	99%
% tested	10%	15%	21%	12%	10%	15%	22%	12%
20% VL testing								
	Probability range tested: 0.23 - 0.64				Probability range tested: 0.21 - 0.74			
PPV	94%	89%	78%	90%	99%	98%	98%	99%
NPV	91%	80%	89%	89%	90%	76%	88%	90%
Sensitivity	52%	41%	60%	39%	44%	36%	48%	35%
Specificity	99%	98%	95%	99%	100%	100%	100%	100%
% tested	20%	28%	39%	22%	20%	29%	39%	22%
40% VL testing								
	Probability range tested: 0.15 - 0.78				Probability range tested: 0.15 - 0.75			
PPV	99%	99%	95%	99%	99%	99%	98%	100%
NPV	94%	86%	93%	93%	94%	82%	92%	93%
Sensitivity	70%	61%	76%	61%	65%	56%	68%	58%
Specificity	100%	100%	99%	100%	100%	100%	100%	100%
% tested	40%	50%	64%	46%	40%	49%	61%	43%

Table S6. Accuracy of prediction of virologic failure in derivation and validation cohorts. Results are shown for no VL testing and for the testing of 20% or 40% of patients, using optimal rules for the range of patients tested based on the predicted probability of virologic failure.

Results from analysis where all patients contributed the same weight, independent of the number of measurements during follow-up.

	Model 1 (Current and baseline CD4 count)				Model 2 (Current CD4 count and CD4 count measured 3 months previously)			
	South Africa (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)	**South Africa** (derivation)	**South Africa** (validation)	**Zambia** (validation)	**Asia-Pacific** (validation)
0% VL testing								
	Probability cut off: 0.38				Probability cut off: 0.36			
PPV	65%	50%	45%	50%	59%	51%	38%	44%
NPV	87%	81%	86%	89%	87%	79%	86%	89%
Sensitivity	33%	24%	46%	30%	24%	18%	31%	22%
Specificity	96%	93%	85%	95%	97%	95%	89%	96%
10% VL testing								
	Probability range tested: 0.29 - 0.55				Probability range tested: 0.26 - 0.56			
PPV	88%	81%	73%	79%	89%	83%	78%	82%
NPV	89%	84%	90%	91%	89%	81%	89%	91%
Sensitivity	43%	34%	58%	39%	36%	28%	45%	32%
Specificity	99%	98%	94%	98%	99%	98%	97%	99%
% tested	10%	14%	22%	12%	10%	13%	22%	12%
20% VL testing								
	Probability range tested: 0.22 - 0.64				Probability range tested: 0.2 - 0.67			
PPV	95%	92%	85%	90%	96%	93%	91%	94%
NPV	91%	87%	93%	93%	91%	84%	91%	92%
Sensitivity	55%	47%	71%	52%	49%	40%	56%	45%
Specificity	99%	99%	97%	99%	100%	99%	99%	100%
% tested	20%	26%	40%	24%	20%	24%	39%	24%
40% VL testing								
	Probability range tested: 0.14 - 0.76				Probability range tested: 0.13 - 0.94			
PPV	99%	98%	94%	97%	100%	100%	100%	100%
NPV	94%	91%	96%	95%	94%	88%	94%	95%
Sensitivity	73%	66%	85%	71%	67%	59%	73%	62%
Specificity	100%	100%	99%	100%	100%	100%	100%	100%
% tested	40%	49%	65%	45%	40%	45%	62%	44%