SUPPLEMENT DIGITAL CONTENT

Statistical Analysis - additional information

Time windows used for VI calculation were (-180, 0) days for HI, (-90, 90) days for the 6 month visit, and (-90,180) days for the one year visit. The VI value at a particular time point was computed based on the visit closest to the time point within the corresponding window.

To evaluate the combination of VI values at the three time points, Cox models with predictors combining information from VI's at the three time points were used to evaluate improvement in the predictive value of the VI at one year after HI. Notice that our goal here was to assess the additional predictive power obtained using the patient's history as captured by prior VIs, not to consider the VI as a prospective time-varying predictor. In particular, risk groups were defined in terms of the VI values at the three time points.

Discrimination was evaluated using both the time to death outcome, and the five year mortality outcome. Given the small number of events, the VI was stratified in tertiles. The tertile risk groups were defined by VI values of 0-13, 14-23, 24-118 at HI, 0-6, 7-16, 17-110 at 6M, and 0-6, 7-16, 17-113 at 1Y.

Combining VI – additional information

However, to assess dependency on the number of risk groups and cutoff, we also evaluated the continuous NRI [11], the "category-less" NRI. The net positive reclassification was 0.638 among cases, and -0.308 among controls, for a net continuous reclassification improvement of 0.330 (95% CI (0.080, 0.491)).

Figure S3. Separate histograms of the reduced VI, a weighted combination of age, CD4 and HIV viral load only, for subjects with observed index and for subjects with missing VACS index value.

Multiple Imputations were used to assess the effect of missing VACS index values on the NRI. The result of 200 imputations is depicted in Figure S4.

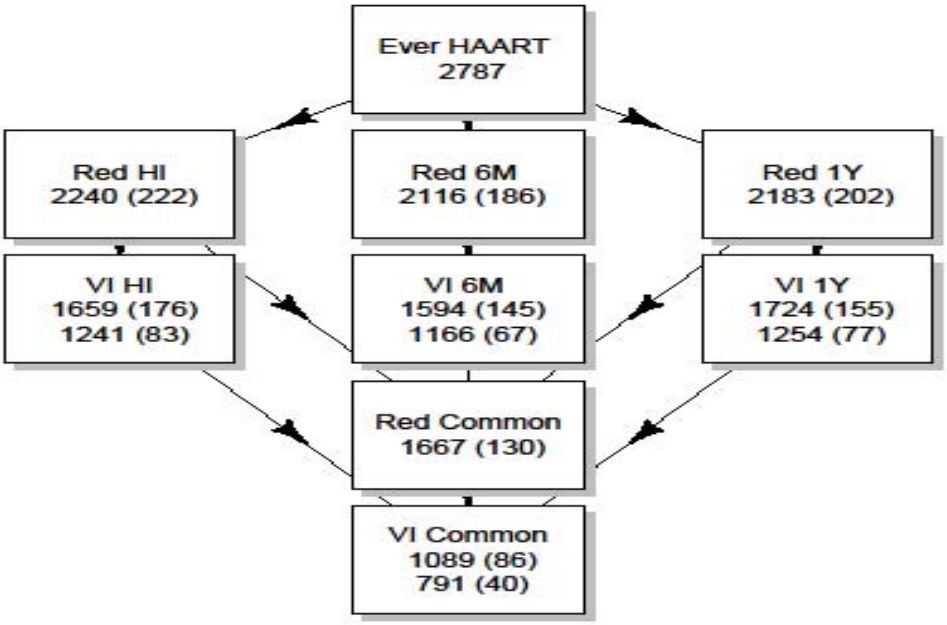

Figure S1. Data flow.

Correlation: Harrell'c and Sommers' D correlation across age quartiles.

Table S1. Harrell'c and Sommers' D correlation across age quartiles.

	Age quartiles			
	18-28.6	28.6-34	34-39.8	>39.8
Harrell' c	0.67	0.74	0.78	0.69
Sommers' D	0.34	0.47	0.57	0.38
Harrell' c	0.62	0.79	0.78	0.76
Sommers' D	0.24	0.58	0.56	0.52
Harrell' c	0.69	0.80	0.83	0.73
Sommers' D	0.38	0.60	0.65	0.46

Discrimination

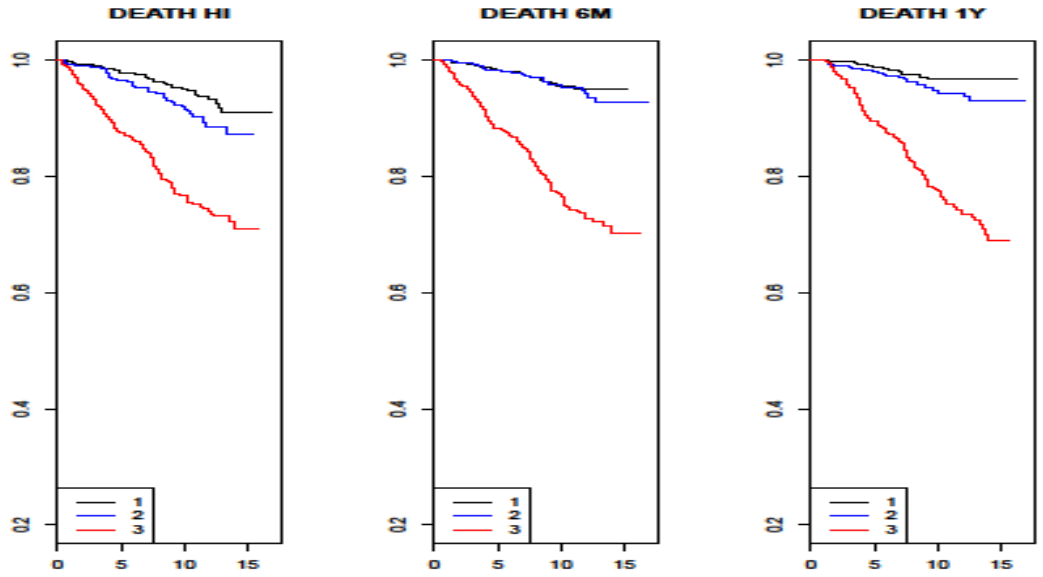

Figure S2. Kaplan-Meier curves for time-to-death at the three time points, stratified by tertiles of the VI.

Calibration: compare the observed and predicted five year mortality (VACS Index at 1Y is the timepoint of primary interest).

Table S2. Observed versus predicted mortality, stratified by VACS Index tertiles.

	HI		6M		1Y	
	Predicted	Observed	Predicted	Observed	Predicted	Observed
I	3.92%	2.48%	2.26%	2.30%	2.99%	1.85%
II	6.88%	3.93%	4.24%	2.13%	4.22%	3.05%
III	24.8%	13.6%	16.4%	12.5%	15.7%	13.3%

Missing values

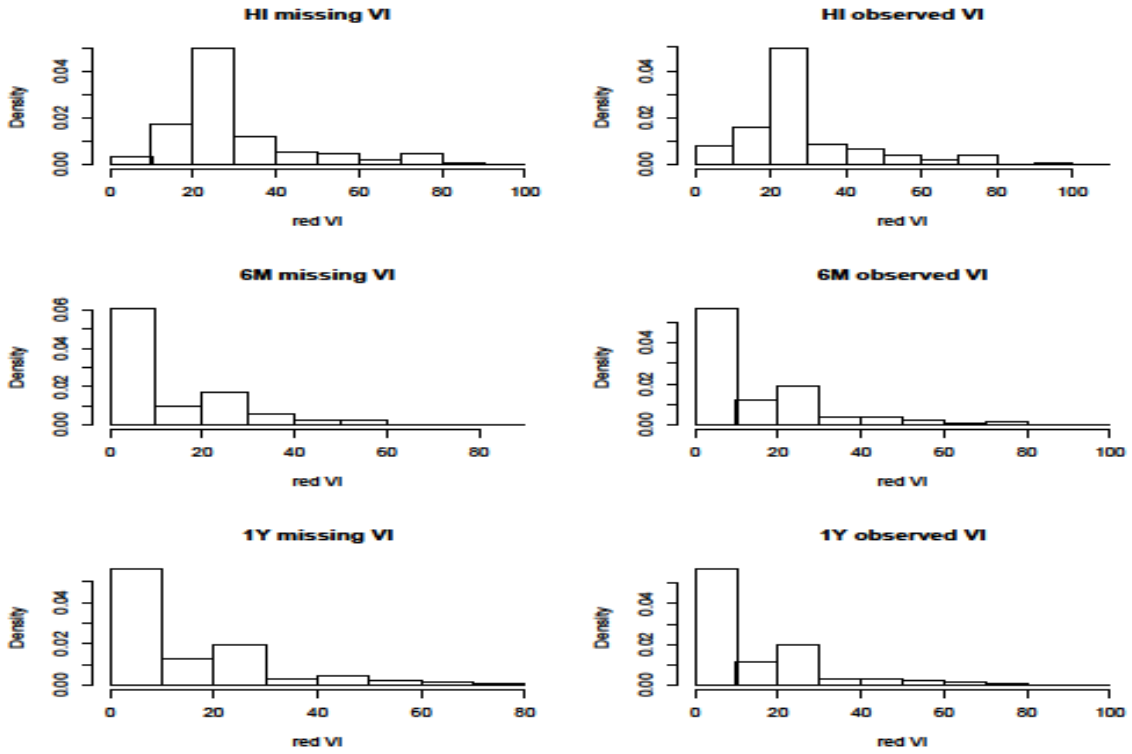

Figure S3. Histograms for the reduced VI for subjects with observed full index and for subjects with missing VI

Combining VI's

Table S3. Observed counts (number of deaths) and proportions for the five year mortality stratified by VACS Index tertiles(cells corresponding to *lower risk history* in green). HI=HAART initiation; for tertiles, 1=lowest, 2=middle, 3=highest.

		VACS Index at 1Y Tertile 1			VACS Index at 1Y Tertile 2			VACS Index at 1Y Tertile 3		
		VACS Index at 6M			VACS Index at 6M			VACS Index at 6M		
	Tertile	1	2	3	1	2	*3*	1	2	3

VACS Index at HI	1	442(7) 0.016	35(2) 0.057	0(0) NA	35(0) 0	39(4) 0.103	1(1) 1	1(1) 1	1(0) 0	1(0) 0
	2	52(1) 0.019	16(0) 0	0(0) NA	7(0) 0	59(6) 0.102	4(0) 0	0(0) NA	10(2) 0.2	10(3) 0.3
	3	2(1) 0.5	5(0) 0	3(0) 0	1(0) 0	22(0) 0	13(2) 0.154	0(0) NA	7(0) 0	25(10) 0.4

Table S4. Observed counts, number of deaths, and proportion stratified by VI tertiles at the three time points.

		VI at 1Y								
		1			2			3		
		VI at 6M			VI at 6M			VI at 6M		
	Tertile	1	2	3	1	2	3	1	2	3
VI at HI	1	653(15) 0.023	40(4) 0.1	0(0) NA	38(1) 0.026	43(9) 0.209	1(1) 1	1(1) 1	1(0) 0	1(0) 0
	2	96(2) 0.021	18(0) 0	0(0) NA	11(1) 0.091	73(15) 0.205	6(1) 0.167	0(0) NA	12(4) 0.333	10(4) 0.4
	3	3(1) 0.333	6(0) 0	3(0) 0	2(0) 0	25(6) 0.240	13(4) 0.308	0(0) NA	8(1) 0.125	25(16) 0.64

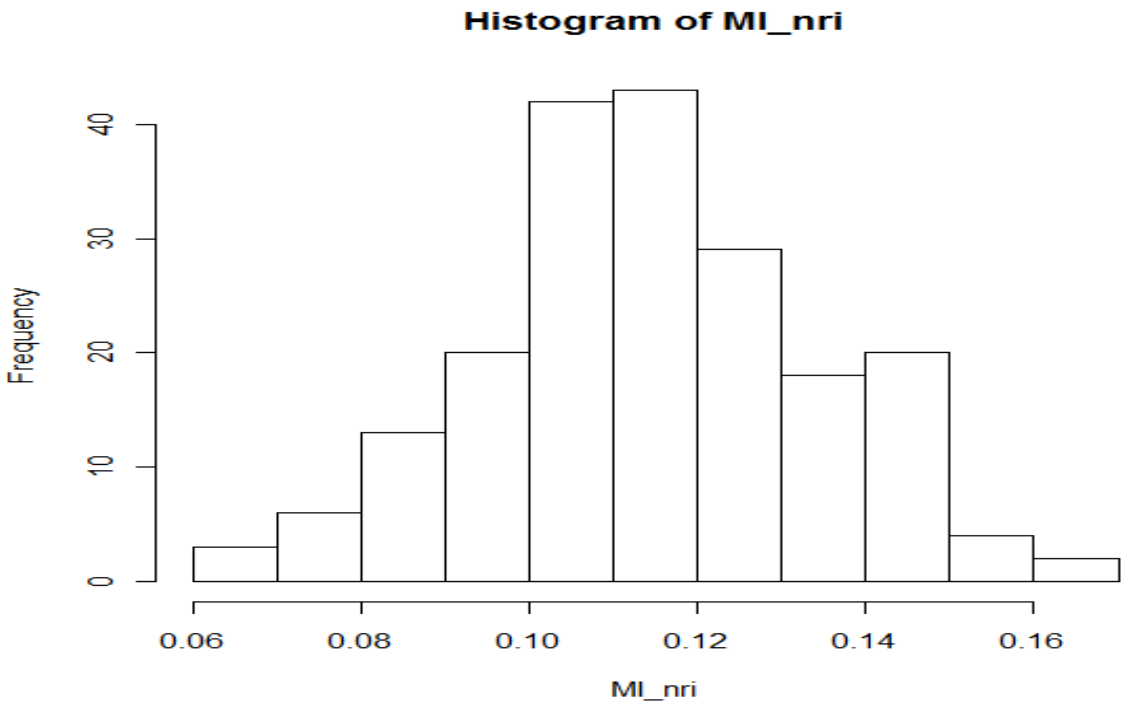

Figure S4. Net reclassification improvement for 200 multiple imputations using the reduced VACS index to impute the missing VACS index values.

We also investigated the relative stability of cohort enrollees over the follow-up. To evaluate this with regard to the VACS Index analyses, we compared the VACS Index values for the 1659 subjects included in the study, stratified by:

i) cohort entry date (before 1/1/1999, 1/1/1999 - 12/31/2003, 1/1/2004 and after), and
ii) HAART initiation date (same date ranges).

Boxplots for the index values at HI appear in Figure S5 below, showing they are quite similar over time in all strata.

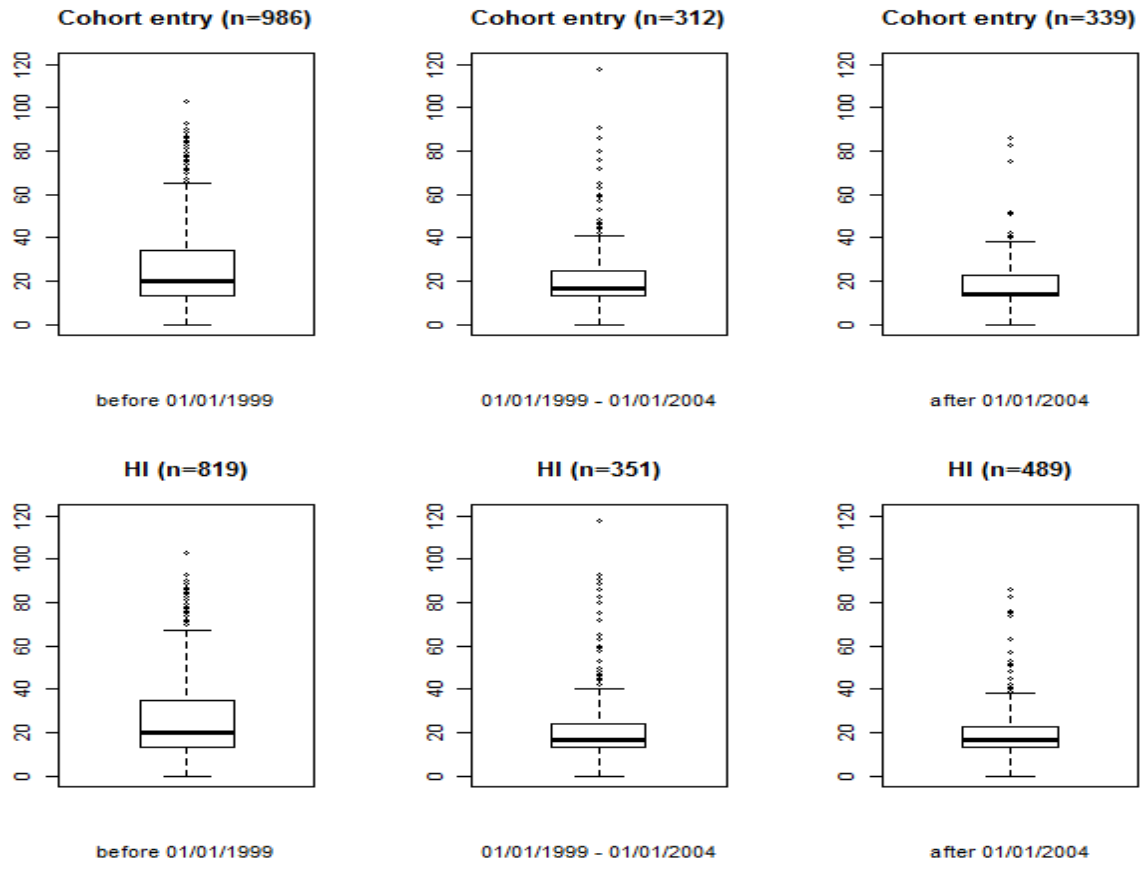

Figure S5. Boxplots for the VACS Index values stratified by cohort entry (first row) and HAART initiation (second row).

Combining VI – additional analyses

Principal components model

Using the VACS Index values at the three time points, construct the principal components (PCs), and then run the Cox model on the PCs.

```
> PC
$values
[1] 572.40965  88.06325  41.20661
$vectors
                PC1          PC2           PC3
[HI]  -0.619212      0.7626310     0.1870039
[6M] -0.572761    -0.2757608    -0.7719460
[1Y] -0.537141     -0.5851069     0.6075598
```

The columns correspond to the coefficients of the first, second and third PC, while the rows to the three time points. For example,

PC1 = -0.619*VACS INDEX_HI - 0.572*VACS INDEX_6M - 0.537*VACS INDEX_1Y

PC2 = 0.762*VACS INDEX_HI - 0.275*VACS INDEX_6M - 0.585*VACS INDEX_1Y

Qualitatively, this means PC1 is approximately an average of the three values, while PC2 is approximately the difference between VACS Index at HI versus 6M and 1Y.

The first PC explains 81.57% of the variability in the VACS Indexes at the three time points, the first two PCs explain 94.12% of the variability.

Ran the Cox model on the PCs, and got the model with the best AIC:

```
              coef  exp(coef)   se(coef)         z       Pr(>|z|)
PC1    -0.026902  0.973456  0.002469   -10.89     < 2e-16 ***
PC2    -0.024102  0.976186  0.007659   - 3.14       0.0016 **
---
```

The model selected includes the first two PCs. Qualitatively, the coefficient for PC2 is negative, meaning that if the VACS index decreases, then the PC2 will tend to be positive, and therefore results in a lower risk of death.

History variable model

Cutoffs for the first and second tertiles of the VI at HI among the subjects who died were 23 and 48. The VI medians at the three time points were 17, 10, and 10.

Taking HI as the baseline (reference) time point, the risk groups at 6M and 1Y were normalized by adjusting for the VI median reduction at the corresponding time point, which resulted in the following cutoff values: (23,48) at HI, (16,41) at 6M, and (16,41) at 1Y.

We used the standard 3-category risk stratification, with cutoffs defined by the tertiles of the predicted mortality among cases in the model only including VI at 1Y.

Taking into account the information in Table2, and Table S3, a new binary variable, “history”, was created. Subjects who started and remained in the lowest tertile or started in the second and moved to the lowest, were categorized as having a lower risk history while subjects who did not meet this definition were categorized as having a higher risk history. The variable *history* was independently associated with all-cause mortality (unadjusted log-rank $p<0.05$, adjusted for VI at 1Y HR=3.93, $p<0.05$), and five year mortality (unadjusted OR=8.23, $p<0.001$, and adjusted for VI at 1Y OR=2.86, $p<0.05$). Inverse probability weighting was used in the five year mortality logistic regression model to account for censoring within the first five years.