

eAppendix

1. *Serious Non-AIDS Events and Data Collection*

Serious non-AIDS events included the following: coronary artery disease, coronary arteriosclerosis, coronary atherosclerosis, acute myocardial infarction, cirrhosis of liver, esophageal varices, hepatic failure, intestinal adenocarcinoma, malignant neoplasm of colon, small cell carcinoma, melanoma of skin, squamous cell carcinoma of anus, end stage renal disease, hepatic coma, cerebral artery occlusion with cerebral infarction, cerebral hemorrhage, hepatic encephalopathy, hepatocellular carcinoma, squamous cell carcinoma keratinizing, and subarachnoid hemorrhage. These were all noted as established diagnoses by ICD9 in the electronic medical record; when onset dates were not in the electronic medical record, the patient's records were reviewed and the appropriate onset date was recorded.

Clinical data were entered into an electronic medical record by medical providers at the time of the patient encounter by automated data upload or by clinic personnel. Laboratory data and all ART use (including dates) were validated by systematic chart review. All events were reviewed and confirmed by study investigators. Information on death was obtained from a registry maintained by the clinic and regular checks of the national Social Security Death Index.

2. *Formula for Utility 2*

$y=0$ if subject is deceased by month k , $y=0.65$ if subject has cytomegalovirus or a fungal ADE by month k , $y=0.61$ if subject has pneumocystis jirovecii pneumonia by month k , $y=0.56$ if subject has any other ADE or NADE by month k , and if subject is asymptomatic at month k ,

$y=0.78+0.0004\times CD4$ if $CD4<75$, $0.75+0.0008\times CD4$ if $75\leq CD4<150$, $0.765+0.0007\times CD4$ if $150\leq CD4<250$, $0.93+0.00004\times CD4$ if $250\leq CD4<500$, or 0.95 if $CD4\geq 500$. This utility is based on a published quality of life metric given by Freedberg and colleagues.¹

3. Regimen Rules

The regimen rule is defined as *start HAART within 3 months of first CD4 measurement dropping below x*. We specified the set of all treatment rules compatible with each patient's CD4 and HAART initiation history. This was done using the following algorithm:

I. For patients with more than 3 months of follow-up after study entry (date of first $CD4<750$), consider the following timeline:



where

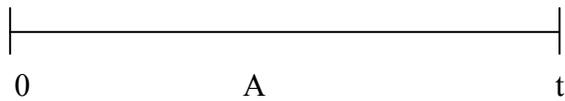
- B = minimum CD4 during the time period $[0, t-3)$ months
- A = minimum CD4 during the time period $[t-3, t]$ months
- t = earliest of 1) date of HAART initiation, 2) date of ADE or death, 3) date of last visit, or 4) k months.

Regimen Rules:

1. If $B > A \geq 200$ and $t = \text{date of HAART initiation}$, then $x = B, B-1, \dots, A+1$
2. If $B > A \geq 200$ and $t \neq \text{date of HAART initiation}$, then $x = B, B-1, \dots, 201$
3. If $A \geq B > 200$ and $t = \text{date of HAART initiation}$ then x is missing and artificially censor at date of HAART initiation
4. If $A \geq B > 200$ and $t \neq \text{date of HAART initiation}$ then $x = B, B-1, \dots, 201$
5. If $B > 200 > A$, then $x = B, B-1, \dots, 201$
6. If $B \leq 200$, then x is missing and artificially censor 3 months after date of first $CD4 \leq 200$
7. If A is missing and $B > 200$ and $t = \text{date of HAART initiation}$, then x is missing and censor at date of HAART initiation
8. If A is missing and $B > 200$ and $t \neq \text{date of HAART initiation}$, then $x = B, B-1, \dots, 201$

II. For patients with no more than 3 months of follow-up after study entry, consider the

following timeline:



where

A = minimum CD4 during the time period [0,t] months
t is defined as before

Regimen Rules:

9. If $A \geq 200$ and $t = \text{date of HAART initiation}$, then $x = 750, 749, \dots, A+1$
10. If $A \geq 200$ and $t \neq \text{date of HAART initiation}$, then $x = 750, 749, \dots, 201$
11. If $A < 200$, then $x = 750, 749, \dots, 201$

The rules used to determine compatible regimens for the patients given in Table 1 of the manuscript were the following: patient A- rule 1; patient B- rule 3; patient C- rule 7; patient D- rule 6; patient E- rule 5; patient F- rule 9; patient G- rule 9; patient H- rule 11; patient I- rule 10; patient J- rule 8; patient K- rule 2; patient L- rule 4.

It should be noted that for computing regimen rules we measured time using dates, and that 3 months was converted to $365.25 \times 3 / 12 = 91.3125$ days. Therefore, it was impossible for the CD4 measurement to be taken exactly 3 months before t. However, as pointed out by a referee, if A was measured exactly 3 months before t (which is possible had we defined 3 months as, for example, 90 days), then cases 2, 5, and 10 given above for assigning regimen rules should be modified so that for cases 2 and 10, $x = A, A-1, \dots, 201$; and for case 5, x is missing and artificially censor at time t. If A was not measured exactly 3 months before t, then the rules are correct as given above.

4. Details of Models used for Inverse Probability Weights

The probability of starting treatment was modeled using logistic regression with the covariates months from study entry (expanded using restricted cubic splines (RCS) with 3, 4, or 5 knots for $k=6, 12, \text{ or } >12$ months, respectively), race (African American/non-African American), injection drug use as HIV risk factor (IDU), sex, age at study entry, most recent CD4 (expanded using RCS with 3 knots), CD4 percent, square-root-transformed time since most recent CD4 measurement, log-transformed HIV-1 RNA, square-root-transformed time since most recent HIV-1 RNA measurement, and square-root-transformed time in care.

The probability of being lost to follow-up was modeled using logistic regression and the covariates months from study entry (expanded using RCS as detailed above), race, IDU, sex, age at study entry, most recent CD4 (expanded using RCS with 3 knots), CD4 percent, log-transformed HIV-1 RNA, square-root-transformed time in care, and time on HAART. Stabilized weights were computed using a second logistic model only including the covariates months from study entry (expanded using RCS as before), race, IDU, sex, and age at study entry. The probability of not being followed at k months was modeled using logistic regression with a single observation per subject and with study entry values for the same covariates included in the lost to follow-up model, except not including months from study entry and time on HAART.

5. Estimating the Maximum Predicted Value

The optimal rule for starting HAART was estimated by regressing utility scores (y) on regimen rules (x) using weighted least squares regression allowing for a flexible/non-linear relationship between x and y , and then finding the value of x which maximized the expected value of y .²

Such a procedure tends to lead to a disproportionate number of estimates of the optimal x at the boundary of the allowed range of x , as demonstrated in the following simulation.

One thousand values of x were randomly drawn from a uniform distribution with bounds of 200 and 750, and 1000 values of y were randomly drawn from an independent standard normal distribution (mean zero and variance 1). Using least squares regression, we regressed y on x with x expanded using a polynomial of degree 5, i.e., $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5$.

From the estimated parameters we found the optimal x : the value of x in [200, 750] which maximized the expectation of y . This estimate was saved, and we repeated the procedure 1000 times. The left side of eFigure 1 is a histogram showing the distribution of the resulting 1000 estimates of the optimal x . In this simulation, the optimal x was estimated to be 200 in 27.8% of simulation replications and 750 in 27.3% of simulation replications. We performed the same procedure regressing y on x using least squares regression with x expanded using restricted cubic splines with 6 knots set to the default locations of R's Design package. The right side of eFigure 1 is a histogram showing the distribution of the estimates of the optimal x using this slightly different model; the shape is similar to the distribution using polynomials, although there is an interesting pattern of high/low frequency estimates between 200 and 750, presumably driven by the knot locations. In this simulation, the optimal x was estimated to be 200 in 22.7% of simulation replications and 750 in 22.6%.

In short, in these simulations where data were generated such that x and y were independent, the estimated optimal x was not uniform, but was disproportionately 200 or 750, the minimum and maximum possible values. The choice of distributions and the inclusion of weights have no

impact on this general conclusion. Even if there is a dependent relationship between x and y (as we suspect in our study) and if the optimal x is some value inside 200 and 750, we still expect to see an inflated rate of estimates at the boundaries 200 and 750.

6. Analysis of EFV-Based Regimens

HAART regimens in both the ART-CC and NA-ACCORD studies were predominantly non-ritonavir-boosted protease inhibitor regimens which are currently rarely used because of inferior potency.^{3,4,5,6,7} In order to assess the model performance with more contemporary HAART regimens, we performed secondary analyses restricted to modern, potent efavirenz (EFV)-based regimens. These analyses were identical to those described in the main text except in these analyses we artificially censored individuals who started a regimen not containing EFV. As with our other censored individuals, we incorporated inverse probability weights to address potential bias arising from this additional artificial censoring.

Of those who initiated HAART during follow-up, 31% (195/621) started an EFV-based regimen. The number of patients who were censored for starting a regimen without EFV is given in eTable 1. Results for these secondary analyses are given in eFigure 2. To maximize health as defined by utility 1, the optimal rule for starting an EFV-based regimen was to start within 3 months of CD4 first dropping below 520 (95% CI=493-550), 509 (95% CI=428-559), 464 (95% CI=418-554), and 485 (95% CI=413-577) for $k=6, 12, 24,$ and 36 months, respectively. The optimal time to initiate HAART based on maximizing utility 2 (quality-of-life) at $k=6, 12, 24,$ and 36 months was the first time CD4 dropped below 286 (95% CI=201-421), 320 (95% CI=201-380), 337 (95% CI=263-453), and 336 (95% CI=201-512), respectively.

Surprisingly, our estimated CD4 thresholds for starting HAART were slightly lower (although still within the 95% confidence limits) when only considering EFV-based regimens, and confidence intervals were generally narrower despite the fewer number of non-missing outcomes in these analyses. This suggests that trends in the health metric score across treatment rules were less variable for patients with outcomes in these analyses.

7. Analyses where Candidate Rules were Limited to CD4=201-500

The NA-ACCORD study suggested the optimal CD4 level for starting HAART was greater than 500. Hence, candidate rules for our analyses were *start HAART within 3 months of first CD4 measured below x* , with $x=201, \dots, 750$. However, we had initially planned to do analyses with candidate rules of $x=201, \dots, 500$. Here we present results of analyses identical to those presented in the Results section of the manuscript except the candidate rules were limited to $x=201, \dots, 500$. Individuals eligible for this secondary analysis were those who presented to the clinic antiretroviral-naïve and had at least one CD4 in the range 200-499; a total of 798 patients met inclusion criteria.

To maximize health as defined by utility 1, the optimal rule was to start HAART within 3 months of first CD4 measured below 428 (95% CI=413-500), 449 (95% CI=402-500), 500 (95% CI=402-500), and 471 (95% CI=410-500) for $k=6, 12, 24,$ and 36 months, respectively. These estimates are roughly consistent with estimates from primary analyses using utility 1 and the complete set of candidate rules.

Using the restricted set of candidate rules, the optimal time to initiate HAART based on maximizing utility 2 (quality-of-life) at $k=6, 12, 24,$ and 36 months was estimated as the first time CD4 was measured below 341 (95% CI=201-396), 355 (95% CI=263-375), 371 (95% CI=266-500), and 384 (95% CI=265-500), respectively. These estimates are consistent with estimates from primary analyses using utility 2 and the complete set of candidate rules.

8. Analyses with Different Utilities

We also performed secondary analyses considering different utilities. In addition to utilities 1 and 2 given above, we also considered the following utilities:

3. Survival: $y=0$ if dead by month k , 1 if alive.
4. ADE-free survival: $y=0$ if dead or ADE by month k , 1 otherwise.
5. ADE/NADE-free survival: $y=0$ if dead, ADE, or NADE by month k , 1 otherwise.
6. CD4-based with regimen changes: Utility 1 if died, otherwise utility 1 multiplied by 0.9^m , where m was the number of regimen changes by month k .
7. QOL with regimen changes: Utility 2 multiplied by 0.9^m , where m was the number of regimen changes by month k .
8. CD4-based without NADE: Utility 1 except ignoring NADE (those with NADE were assigned utility 1 value they would have received depending on any subsequent ADE or their CD4 at month k).
9. QOL without NADE: Utility 2 except ignoring NADE (those with NADE were assigned utility 2 value they would have received depending on any subsequent ADE or their

CD4 at month k).

The estimated optimal rules and confidence intervals for all utilities are given in eTable 2.

Utilities 8 and 9 (CD4-based and QOL without NADE) yielded results very similar to utilities 1 and 2 (CD4-based and QOL including NADE), respectively, suggesting that in our cohort NADE had little impact on the estimated optimal time to initiate HAART. Results using utility 6 (CD4-based with regimen changes) were fairly consistent with results using utility 1 (CD4-based) except the optimal CD4 at which to initiate HAART was generally lower, which is sensible given that utility 6 assigned lower health scores to individuals who had changed regimens. More puzzling is that utility 7 (QOL with regimen changes) estimated the optimal CD4 to initiate HAART was 201 for $k=6, 12, 24,$ and 36 months. This is explained at least in part by the method's tendency to estimate optimal rules at the extremes of the candidate rules (see appendix section titled *Estimating the Maximum Predicted Value*), but we would have expected more variation. The estimates based on utility 3 (survival) were quite variable because there were very few deaths in our cohort; 95% confidence intervals spanned the range of candidate rules. Utilities 4 and 5 (ADE-free survival and ADE/NADE-free survival, respectively) were based on more events, although not many more (see Table 1), and tended to favor initiation of HAART at lower CD4 levels.

All of the utilities except 1, 6, and 8 put considerable weight on deaths or other events, of which there were few in our cohort. We believe that our estimates would be more reliable and consistent in a larger cohort with more events.

9. Pre-HAART Deaths within 3-months of Study Entry

The NA-ACCORD study was criticized because those who died within 6 months of their baseline CD4 measurement without starting HAART were assigned to the deferred group, possibly biasing results.⁸ In our analysis, anyone who died within 3 months of their first CD4 between 200-749 without starting HAART would be compatible with all treatment rules x , and would therefore add no information to the analysis (i.e., each treatment rule would be given the same outcome).

10. Inclusion Criteria

Consider a randomized trial that investigates the rules *start HAART within 3 months of CD4 first measured below $x=201-750$* . Individuals are randomized to a value of x at the date of their first measured CD4 <750 , which is defined as study entry (or time 0). In order to be included in this trial, patients would have to have at least one pre-HAART CD4 measurement in the range 200-749. We will refer to this inclusion criterion as option A. In contrast, one could further restrict the inclusion criterion to only include patients who have a pre-HAART CD4 >749 before their CD4 measurement between 200-749. We'll refer to this as option B.

There are advantages and disadvantages to the two competing inclusion criteria. Since all patients have at least one CD4 measurement greater than 749 with option B, then all patients will start the trial with a similarly high CD4 at time 0 (assuming CD4 measurements are regularly taken when patients' CD4 >749). Therefore, the trial should have good power to detect differences between x because the inclusion criterion has controlled for the variation in CD4 counts at time 0. The disadvantage to such a trial is that it is difficult to identify HIV-positive

individuals who have $CD4 > 749$, as most patients do not know of their infection status until they have lower CD4 levels. Therefore, recruitment with option B as the inclusion criteria will be difficult. Furthermore, since patients rarely present to clinics with $CD4 > 749$, those who do may not be representative of typical HIV-patients, and thus results may not generalize to standard patient populations.

In contrast, option A does not control variation in CD4 counts at time 0. For example, suppose a patient presents to the clinic with $CD4 = 400$, discovers he is eligible for the trial, enrolls, is randomized to $x = 700$, and thus immediately starts HAART. Notice that if this person had been randomized to $x = 450$, he would also have immediately started HAART. Hence, option A's inclusion criterion makes it difficult to distinguish between higher rules, such as $x = 450$ vs $x = 700$, because CD4 counts at time 0 are not controlled to be high. It will be particularly difficult to distinguish between very high rules such as $x = 650$ vs $x = 750$ under option A. (In contrast, the trial will be good at distinguishing between lower rules such as $x = 250$ vs $x = 350$.) It should be noted, however, that such a trial is still valid: As the sample size gets big and hence there are more people whose first CD4 measurements are in the higher range of values, the power to detect true differences between any two rules will go to 100%. However, to detect a true difference between higher rules (e.g. $x = 650$ vs $x = 750$), many more patients would need to be enrolled under option A's inclusion criterion than option B's. The primary advantage to option A's design, therefore, is that recruitment will be much easier. In fact, to achieve the same power as option B, option A may actually permit recruiting from a smaller patient population because of its less stringent inclusion criterion. And because of the weaker inclusion criterion, results under option A will be more applicable to standard patient populations as most patients first present to

care with CD4 in the range 200-749.

The method outlined by Robins and colleagues² and employed here treats observational data as if they came from a randomized trial, and similarly requires setting inclusion criteria. One could choose either option A (pre-HAART CD4 between 200-749) or option B (pre-HAART CD4 between 200-749 and at least one CD4>749). The arguments for either option are similar to those discussed above in the context of a randomized trial. Our primary analyses used option A for practical reasons (so that we would not exclude the vast majority of our patients from the analysis) and because we felt that results under this criterion would be more generalizable. Because we had relatively few patients with high pre-HAART CD4 measurements, our inclusion criterion means that, similar to a randomized trial with this design, it is difficult to distinguish between higher treatment rules (e.g., $x=650$ vs $x=750$). This may be seen in Figure 2D, where the expected utility is very flat at the higher treatment rules. The flat expected utility could be due to a lack of power, or that there truly is little difference between higher values of x , or a combination of both. However, this design is still valid, and there are enough patients so that such a design can distinguish between lower treatment rules (e.g., 250 vs. 350) or between a lower and a higher rule (e.g., 250 vs. 750). For example, a patient whose first measured CD4 is 400 and who starts treatment within 3 months of this measurement will have data compatible with rules $x=401, \dots, 750$. Therefore, this patient provides no information to distinguish between the rules $x=450$ and $x=750$, but he does provide information to distinguish between the rules $x=450$ and $x=350$. By combining similar information from hundreds of other patients whose data is consistent with different rules x , we are able to estimate which rule results in the highest expected utility.

We also performed a secondary analysis utilizing option B's inclusion criteria. Because we have so few patients with $CD4 > 749$, in this analysis we only considered the rules $x=201-500$ and the inclusion criteria were a pre-HAART $CD4 > 499$ followed by a pre-HAART $CD4$ between 200 and 499. Results of this analysis are presented below in section 11.

11. Analyses restricted to those with pre-HAART $CD4 \geq 500$

We performed an additional analysis which was limited to persons who had a measured $CD4 \geq 500$ before initiating HAART. Study entry in this analysis was defined as the date of the first $CD4$ measured between 200-499. We considered the candidate rules *start HAART within 3 months of $CD4$ first measured below x* , with x from 201-500. A total of 83 patients met inclusion criteria.

To maximize health as defined by utility 1, the optimal rule was to *start HAAART within 3 months of first $CD4$ measured below 355 (95% CI=249-500), 404 (95% CI=278-500), 430 (95% CI=275-500), and 435 (95% CI=285-500) for $k=6, 12, 24,$ and 36 months, respectively. These estimates are slightly lower, but comparable to the estimates using utility 1 which only considered the candidate rules $x=201-500$, but which did not require a pre-HAART $CD4 \geq 500$ for inclusion (given in section 7 of Appendix).*

The optimal time to initiate HAART based on maximizing utility 2 (QOL) at $k=6, 12, 24,$ and 36 months was estimated as the first time $CD4$ was measured below 355 (95% CI=201-422), 286 (95% CI=201-500), 260 (95% CI=242-500), and 383 (95% CI=245-500), respectively.

Confidence intervals for these analyses largely overlapped with their counterparts that did not require a pre-HAART CD4 ≥ 500 for inclusion (see section 7 of Appendix).

Appendix References

1. Freedberg KA, Scharfstein JA, Seage GR III, et al. The cost-effectiveness of preventing AIDS-related opportunistic infections. *JAMA*. 1998;279(2):130-136.
2. Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. *Stat Med*. 2008;27(23):4678-4721.
3. 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(5):555-570.
4. Willig JH, Abrams S, Westfall AO, et al. Increased regimen durability in the era of once-daily fixed-dose combination antiretroviral therapy. *AIDS*. 2008;22(15):1951-1960.
5. Shafer RW, Smeaton LM, Robbins GK, et al. Comparison of four-drug regimens and pairs of sequential three-drug regimens as initial therapy for HIV-1 infection. *N Engl J Med*. 2003; 349(24):2304-2315.
6. Palella FJ Jr, Baker RK, Moorman AC, et al.; HIV Outpatient Study Investigators. Mortality in the highly active antiretroviral therapy era: changing causes of death and disease in the HIV outpatient study. *J Acquir Immune Defic Syndr*. 2006;43(1):27-34.
7. Lau B, Gange SJ, Moore RD. Risk of non-AIDS-related mortality may exceed risk of AIDS-

related mortality among individuals enrolling into care with CD4+ counts greater than 200 cells/mm³. *J Acquir Immune Defic Syndr.* 2007;44(2):179-187.

8. Hernan MA and Robins JM (2009). Early versus deferred antiretroviral therapy for HIV [letter]. *N Eng J Med.* 2009; 361(8):822-823.

eTable 1. Number of patients who died, had an AIDS-defining event (ADE), a non-AIDS defining event (NADE), or were censored according to follow-up period in the analyses censoring those who started a regimen without efavirenz (EFV).

Months of Follow-up (k)	Death	ADE	NADE	Asymptomatic	Censored			
					LTF U	End of study	Artificially censored	Started non- EFV regimen
6	1	10	5	649	71	23	41	234
12	6	13	6	518	103	59	74	255
24	13	17	4	328	151	119	116	286
36	23	21	4	228	180	142	137	299

eTable 2. Estimated optimal CD4 level for initiating HAART based on all utilities (and 95% confidence interval).

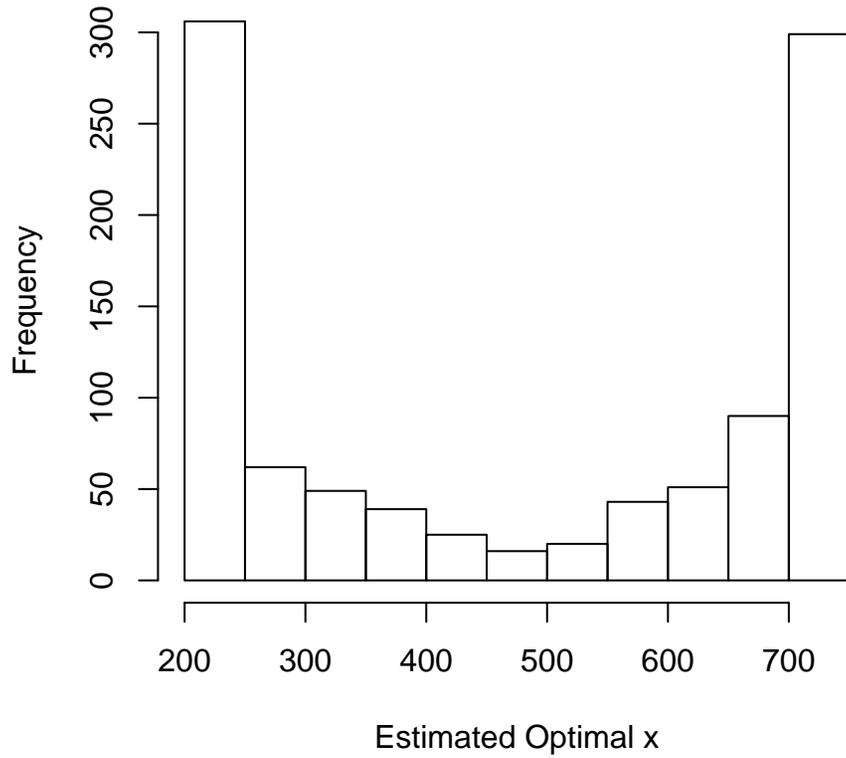
Utility	Months of Follow-up (k)			
	6	12	24	36
1. CD4-based	495 (468-522)	554 (459-750)	489 (427-750)	509 (460-750)
2. QOL	337 (201-442)	354 (288-386)	358 (294-750)	475 (287-750)
3. Survival	201 (201-750)	201 (201-750)	491 (201-750)	563 (201-750)
4. ADE-free survival	201 (201-389)	201 (201-436)	368 (272-750)	450 (227-750)
5. ADE/NADE-free survival	201 (201-400)	327 (201-451)	374 (201-451)	433 (201-750)
6. CD4-based with regimen changes	460 (429-474)	414 (381-448)	406 (364-448)	421 (297-485)
7. QOL with regimen changes	201 (201-201)	201 (201-201)	201 (201-280)	201 (201-347)
8. CD4-based without NADE	501 (472-533)	563 (480-750)	518 (460-750)	531 (495-750)
9. QOL without NADE	319 (201-393)	334 (293-378)	350 (302-750)	750 (302-750)

eFigure 1. The distribution of the estimated optimal x based on 1000 simulation replications.

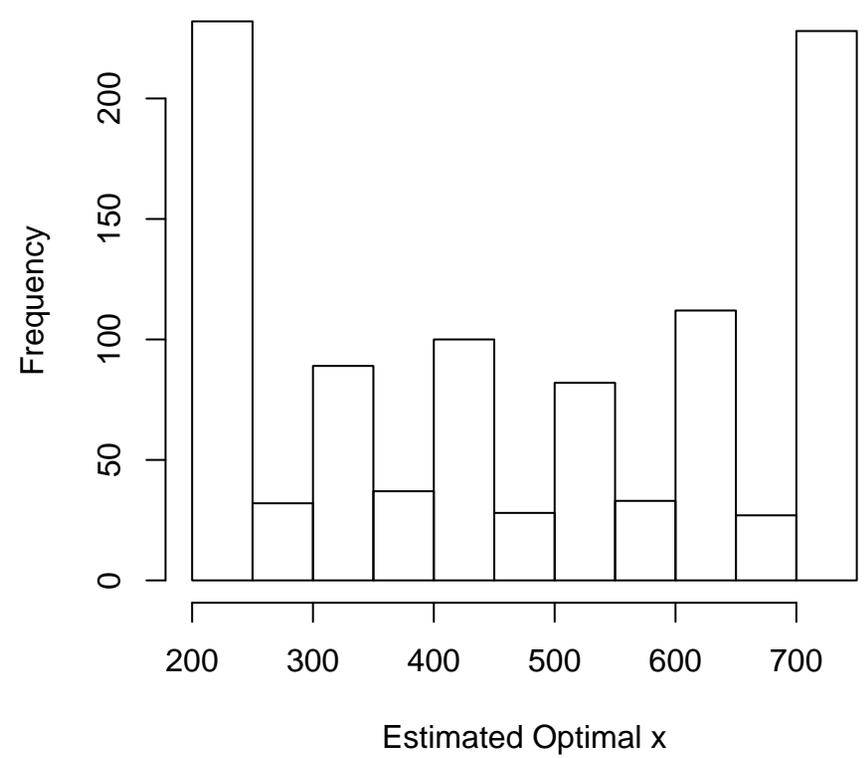
The histogram on the left estimates the optimal x using a regression model with polynomials, the histogram on the right uses restricted cubic splines with 6 knots.

eFigure 2. Estimates and 95% confidence intervals of the optimal CD4 at which to initiate an EFV-containing HAART regimen in order to maximize utilities 1 and 2 at 6, 12, 24, and 36 months after study entry.

Distribution of optimal x using polynomials



Distribution of optimal x using restricted cubic splines



EFV-based regimens

