Published in final edited form as:

J Clin Epidemiol. 2009 July ; 62(7): 729–737. doi:10.1016/j.jclinepi.2008.09.002.

Cross-cohort heterogeneity encountered while validating a model for HIV disease progression among antiretroviral initiators

Bryan E. Shepherd^{a,*}, Timothy R. Sterling^b, Richard D. Moore^c, Stephen P. Raffanti^b, and Todd Hulan^b

^aDepartment of Biostatistics, Vanderbilt University School of Medicine, S2323 MCN, 1161 21st Avenue South, Nashville, TN 37232, USA

^bDepartment of Medicine, Division of Infectious Diseases, Vanderbilt University School of Medicine, Nashville, TN 37203, USA

^cDepartment of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287 USA

Abstract

Objective—To evaluate a model for predicting time to AIDS or death among HIV-infected persons initiating highly active antiretroviral therapy (HAART).

Study Design and Setting—The model was constructed from 1,891 HAART initiators in the Collaborations in HIV Outcomes Research/US (CHORUS) cohort. The model's predictive ability was assessed using internal bootstrap validation techniques and data from 716 HAART initiators at Johns Hopkins HIV Clinical Cohort (JHHCC) in whom HIV disease was, in general, more advanced.

Results—The estimated concordance statistic was 0.632 with the bootstrap method and 0.625 in JHHCC. Mean predicted and observed 3-year AIDS-free survival for JHHCC was 0.76 and 0.73 (95% CI 0.69–0.77), respectively; mean predicted and observed 5-year AIDS-free survival was 0.69 and 0.57 (95% CI 0.52–0.62). Sensitivity analyses showed that the discrepancy between predicted and observed AIDS-free survival after 3 years could be due to differences in lost-to-follow-up rates between cohorts.

Conclusion—The model was fair at using baseline characteristics to order patients' risk of disease progression, but did not accurately predict AIDS-free survival >3 years after HAART initiation. Different variable definitions, patient characteristics, and loss-to-follow-up highlight the challenges of using data from one cohort to predict AIDS-free survival in an independent cohort.

Keywords

AIDS; Antiretroviral therapy; Bootstrap; CD4 lymphocyte percent; Loss-to-follow-up; Validation; Heterogeneity

© 2009 Elsevier Inc. All rights reserved.

* Corresponding Author. Telephone: 615–343–3496. Fax: 615–343–4924. bryan.shepherd@vanderbilt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Observational cohort studies have been vital to our understanding of the natural history of HIV disease progression, and in the identification of factors associated with prognosis after antiretroviral therapy initiation [1-2]. Recognizing the importance of such studies, the National Institute of Allergy and Infectious Diseases at the U.S. National Institutes of Health have sponsored the International Epidemiologic Databases to Evaluate AIDS (IeDEA), a global consortium to combine data and answer questions about the HIV epidemic (www.iedea-hiv.org) [3-4].

Combining data from heterogeneous sites or applying knowledge learned from one site to another can be challenging, as patient characteristics, clinical practice patterns, and even HIV-1 strains can differ between cohorts. Nevertheless, patient characteristics, especially immunologic status, at the initiation of highly active antiretroviral therapy (HAART) have been shown to be associated with subsequent AIDS defining events (ADE) and death [2, 5-11].

Using data from the Collaborations in HIV Outcomes Research/US (CHORUS) cohort, we observed that percent (%CD4) and absolute (aCD4) CD4⁺ lymphocytes at HAART initiation, prior exposure to antiretroviral therapy (ART), and probable route of HIV acquisition (injection drug use [IDU] or non-IDU) were associated with subsequent progression to ADE/death in a multivariable model [12]. Using these patient characteristics, as well as HIV-1 RNA at HAART initiation and demographic features, we created a prognostic model for ADE/death in HIV-infected persons initiating HAART. A calculator that uses this model to predict ADE-free survival probabilities up to seven years after HAART initiation is available at <http://biostat.mc.vanderbilt.edu/HIVSurvivalPrediction>.

The goal of this study was to assess how well our model predicted ADE/death in an independent cohort. First, we used re-sampling techniques to internally estimate our model's predictive ability when applied to independent data. Second, we applied our model to a cohort of HIV-infected persons initiating HAART in Baltimore, MD, the Johns Hopkins HIV Clinical Cohort (JHHCC) Adult HIV Clinic. The JHHCC was specifically chosen in order to evaluate the predictive model on a cohort different from CHORUS in terms of demographics, common routes of infection, and level of immunosuppression at HAART initiation. Finally, we explored possible reasons for the discrepancies between predicted and observed ADE/deaths in the JHHCC.

2. Materials and Methods

2.1. Study Cohorts

Both the CHORUS cohort and the JHHCC have been described in detail previously [13-14]. Briefly, our study included persons from both cohorts who were ≥ 18 years of age and initiated HAART between August 1, 1997 and January 1, 2005, and followed them through August 1, 2005. HAART was defined as at least two nucleoside reverse transcriptase inhibitors (NRTIs) in combination with at least one protease inhibitor (PI) and/or non-NRTI (NNRTI), or at least three NRTIs. To be included in this study, persons had to remain on their first HAART regimen for at least 30 days, have "baseline" (i.e., at time of HAART initiation) demographic data, aCD4, %CD4, and HIV-1 RNA values available within 180 days prior to or 7 days after the date of first HAART, and have a baseline HIV-1 RNA ≥ 5000 copies/mL. HAART was available for virtually all patients in care at all sites during the study period. The study of both cohorts was approved by local Institutional Review Boards, all persons had provided prior written informed consent for use of clinical data, and only aggregated non-identifiable patient data were used in the analysis.

A study event was defined as the first new AIDS defining event (ADE) or death after initiation of the first HAART regimen. ADEs were based on 1993 CDC classification criteria, excluding diagnoses based on $aCD4 < 200$ cells/mm³ [15]. Causes of death in CHORUS and in the JHHCC were determined and categorized as AIDS-related, not AIDS-related, or unknown using pre-specified criteria [16]. At the JHHCC, deaths were determined by ongoing medical record review, and annual surveillance using the national social security index and the National Death Index. Persons who did not have an event were followed until the end of the study period, or until they were lost to follow-up, defined as the day of withdrawal from their cohort or the last clinic visit date for persons who did not have an encounter for nine months for the CHORUS cohort and 18 months for the JHHCC. Analyses were also performed, and are presented in Section 3.3, using the CHORUS definition of lost to follow-up for both cohorts. Secondary analyses added 4.5 months of follow-up time to those lost to follow-up in both cohorts, yielded similar results, and are not presented.

2.2. Model Fitting Procedures using CHORUS data

Our prediction model was originally developed to measure the association between %CD4 and ADE/death after adjusting for aCD4 and other potential confounders [12]. A Cox proportional hazards model was fit to the CHORUS data including the predictor variables baseline %CD4, aCD4, log₁₀-transformed plasma HIV-1 RNA, age, sex, race (non-Hispanic white or non-white by self report), prior ART use, and probable route of HIV infection (IDU vs. other routes). These variables were chosen for inclusion in the model *a priori*, and were included in all candidate models regardless of their statistical significance. Percentage CD4 and aCD4 were square-root transformed to normalize the data. To avoid assuming linearity in the hazard, continuous predictors (%CD4, aCD4, VL, and age) were expanded by fitting restricted cubic splines. For candidate models, each continuous predictor was assigned the same number of knots (0, 3, 4, or 5). The final model used 3 knots (the optimal number of knots based on the Akaike information criteria; knots were located at the default values 10th, 50th, and 90th percentiles). An interaction term between %CD4 and aCD4 was also included in the model. Based on evidence of a possible violation of the proportional hazards assumption [17], the models were stratified on age quintiles, leaving age in the model as a continuous predictor to account for possible residual information [12,18-20]. All analyses were performed using R statistical software (Version 2.3.1; available at <http://www.r-project.org>).

2.3. Measures of the Model's Predictive Ability

The predictive ability of the model was assessed both internally and externally using the following measures chosen *a priori*:

1. Concordance, measured using Harrell's c statistic and defined as the proportion of all patient pairs in which the ordering of the observed times to ADE/death was concordant with the predicted 3-year ADE-free survival probability; $c=0.5$ indicates that a model contains no predictive information. If censoring occurred in such a manner that it was not known who had a longer time to ADE/death, then that pair was not included in the calculation [19].
2. Predictive accuracy, seen by dividing patients into 5 risk categories based on predicted ADE-free survival, allocating patients into groups with approximately equal numbers of events, and comparing mean predicted 3-year ADE-free survival to observed 3-year ADE-free survival based on Kaplan-Meier estimates in each category [18].
3. Shrinkage, defined as the flattening of the plot of predicted vs. observed ADE-free survival away from the 45 degree line; a value of 1 indicates no over-fitting. This was assessed on independent data (within the bootstrap validation or the JHHCC) using

Cox proportional hazards with the observed time to ADE/death regressed on the estimated linear predictor, based on the predictive model [18].

In addition, to further assess our model's predictive ability in the JHHCC, predictive accuracy was computed by comparing the mean predicted survival for the entire JHHCC over the follow-up period with the Kaplan-Meier ADE-free survival curve.

2.4. Bootstrap Procedures for Internal Model Validation

Measures of a model's predictive ability produced by applying the original model to the original data tend to be over-optimistic, i.e. to predict that the model will fit better than it actually does in independent data [21]. The validity of a model applied to independent data can be assessed fairly using re-sampling techniques (bootstrapping) [22]. From the original 1891 rows (patients) in the CHORUS dataset, 1891 rows were randomly sampled with replacement creating bootstrapped data. Model selection techniques identical to those used to create the original model were used to fit a new model to the bootstrapped data. The entire process was repeated to produce a total of 1000 bootstrap replications. Details are described elsewhere [20].

In each bootstrap replication, our measures of the model's predictive ability (measures 1, 2, and 3, discussed above) were computed using the bootstrap model applied to the bootstrapped data. Then the same measures were computed using the bootstrap model applied to the original data. The average difference across bootstrap replications between the measures applying the bootstrap model to the original data versus applying the bootstrap model to the bootstrapped data was used to estimate the measures' optimism. The original model's predictive ability when applied to independent data was then estimated by subtracting a measure's estimated optimism from the original model's measure of predictive ability [18].

2.5. Other Analyses

We tested whether the c-statistic indicated better than chance concordance by performing a permutation test: Each person's follow-up time and event indicator were randomly shuffled and assigned as a pair to another person's predictors. The model was fit following all model selection procedures described above to the permuted data and the c-statistic was computed. This process was repeated 1000 times, and the corresponding p-value was the proportion of permutations resulting in a c-statistic greater than the c-statistic from the original data.

Sensitivity analyses were performed by creating a new model from the CHORUS data in which each person lost to follow-up was simulated to live or have an event (ADE or death), based on a fixed event probability chosen to be the same for all patients lost to follow-up (i.e., not conditional on patient characteristics). If the patient was randomly selected to have an event, the time until their event was then randomly drawn from a uniform distribution between 0 and 1 years after the date of their last visit. The prognostic model was then re-fit to this new CHORUS data and applied to predict mean ADE-free survival in the JHHCC. This process was repeated 100 times, and the average of the 100 repetitions was reported as the mean predicted ADE-free survival for the JHHCC based on the assigned probability of having an ADE/dying if lost to follow-up.

3. Results

3.1. Internal Model Validation

In the CHORUS data, the proportion of all patient pairs in which the ordering of predictions and outcomes were concordant was $c=0.678$. Based on the bootstrap, the expected concordance when applying the model to independent data was $c=0.625$. This suggests that the model is

substantially better than chance at using patient characteristics to order their clinical prognosis ($p < 0.001$).

Figure 1a demonstrates the predictive accuracy of the model. CHORUS participants were divided into 5 categories based on their predicted ADE-free survival, and the figure plots mean predicted 3-year ADE-free survival probability versus the Kaplan-Meier estimate (and 95% confidence interval) for each group. If there were perfect calibration, then the two values would be equal (i.e., fall on the dashed line). The observed ADE-free survivals for given predicted ADE-free survival probabilities when the model is applied to independent data based on the bootstrap are shown as 'X' in the plot. Based on the bootstrap, for individuals with high or low predicted ADE-free survival probabilities, the model is expected to over-estimate or under-estimate, respectively, the true ADE-free survival probability. This suggests that there is some over-fitting, which is further supported by a shrinkage estimate based on the bootstrap of 0.60.

To summarize, based on bootstrap methods of validation, the prognosis model is expected to have fair agreement with observed ADE-free survival when applied to independent data, but is expected to slightly over-estimate ADE-free survival probability for individuals predicted to have a high ADE-free survival probability and to underestimate ADE-free survival for those predicted to have a low ADE-free survival probability.

3.2. Validation Using an Independent Cohort

We next applied our model to predict ADE-free survival probabilities for persons initiating HAART in the JHHCC. Table 1 shows characteristics of CHORUS and the JHHCC. The JHHCC included fewer persons than CHORUS (716 versus 1891), and had shorter follow-up time (median of 2.7 years versus 4.6 years; $p < 0.0001$). Persons initiating HAART at JHHCC had more advanced HIV disease as indicated by lower aCD4 (median 133 vs. 240) and %CD4 (median of 12 vs. 16), and higher plasma HIV-1 RNA (median $\log_{10}$ copies/mL = 4.9 vs. 4.7) ($p < 0.0001$ for each). Persons in the JHHCC were also more likely to be female (35% vs. 10%), non-white (82% vs. 28%), ART naïve before initiation HAART (60% vs. 47%), and to have IDU as a potential route of infection (46% vs. 6%) ($p < 0.0001$ for each).

ADE-free survival in the JHHCC was lower than that observed in CHORUS: 38% of HAART initiators had an event during study follow-up, compared to 25% in CHORUS. Interestingly, the percentage of individuals with an ADE was fairly similar between cohorts (15% in CHORUS vs. 12% in JHHCC). However, 26% of individuals in JHHCC died, compared to 10% in CHORUS. It is worth noting that the causes of death were similar between the two cohorts; in CHORUS, 52% of deaths were AIDS related, 22% not AIDS related, and 26% unknown; whereas in the JHHCC, 53% of deaths were AIDS related, 20% were not AIDS related, and 26% were unknown.

We evaluated our model's predictive ability on the JHHCC by computing the same statistics discussed above. The estimated concordance was 0.632, indicating that our model correctly ordered HAART initiators according to their ADE-free survival 63.2% of the time. This concordance was very similar to that predicted based on the bootstrap validation ($c = 0.625$).

Figure 1b shows the mean predicted 3-year ADE-free survival based on the CHORUS model applied to the JHHCC data, plotted against the actual 3-year ADE-free survival in the JHHCC (Kaplan-Meier estimates and 95% confidence intervals). This plot is very similar to what was expected based on the bootstrap validation (Figure 1a), with those predicted to have very low and very high 3-year ADE-free survival tending to have slightly higher and slightly lower observed ADE-free survival, respectively. Shrinkage in the JHHCC was estimated as 0.58, similar to the bootstrap validation estimate, 0.60, and consistent with Figures 1a and 1b.

Figure 2 shows the estimated ADE-free survival for CHORUS and for JHHCC compared with the mean predicted ADE-free survival in the two arms. The Kaplan-Meier estimate of ADE-free survival probability in the JHHCC was greater than predicted during the first year after HAART initiation, and then much less than predicted more than 3–4 years after HAART initiation. In the JHHCC, the Kaplan-Meier estimates (95% CI) for ADE-free survival at 3, 4, and 5 years were 0.73 (0.69, 0.77), 0.64 (0.60, 0.68), and 0.57 (0.52, 0.62), respectively, whereas the mean predicted ADE-free survival probabilities at those times were 0.76, 0.72, and 0.69. Kaplan-Meier estimates beyond 5 years of follow-up in the JHHCC were quite low, but based on few patients.

To summarize, as expected based on our internal validation, our prognostic model had fair concordance with observed ADE-free survival when applied to independent data from JHHCC. The model slightly over-estimated the 3-year ADE-free survival probability for those predicted to have a high ADE-free survival based on their characteristics at HAART initiation, and slightly under-estimated the 3-year ADE-free survival for those predicted to have low ADE-free survival. However, Figure 2 demonstrates that our model was poor at predicting ADE-free survival for patients 4 or more years after HAART initiation.

3.3. Attempting to Reconcile Predicted and Observed ADE-free Survival in the JHHCC

We explored reasons why the predicted and observed ADE-free survival differed >3 years after HAART initiation in the JHHCC. First, we added new variables to our predictive model and re-fit it to the CHORUS cohort, in hopes that the inclusion of these additional variables might result in a predictive model that would remove some of the difference between the mean predicted and observed ADE-free survival when applied to the JHHCC. A model which included first HAART regimen in addition to all other covariates did not improve the fit (Table 2).

In a recent study, the ART-CC found that the incidence of death for IDUs was roughly constant over 6 years, whereas the incidence of death for non-IDUs dropped sharply during the first year and then remained much lower than that of IDUs [23]. Therefore, we hypothesized that a reason for the difference in ADE-free survival between the two cohorts could be due to the difference between incidence rates (baseline hazards) between IDUs and non-IDUs, particularly because in JHHCC 46% were IDUs versus 6% in CHORUS. Although IDU was included and found to be a significant risk factor in the original model fit to the CHORUS data, we did not stratify by (fit a separate baseline hazard for) probable route of infection because there was little evidence to suggest that the IDU hazard ratio changed with time ($p=0.6$). However, we considered the possibility that this result was due to insufficient power and re-fit the model to CHORUS data, stratifying over IDUs in addition to age categories. This did not account for the difference in observed and predicted ADE-free survival (Table 2). As seen in Figure 3, although the incidence rates were clearly non-proportional between cohorts, they were roughly parallel for IDUs and non-IDUs within each cohort.

We also hypothesized that differences in adherence rates could explain the difference between ADE-free survival in CHORUS and JHHCC. Note that a measure of adherence does not belong in a prognostic model, as adherence is unknown at the time of HAART initiation. This model was for exploratory purposes; including additional *baseline* covariates which help predict adherence would be appropriate for our prognostic model. Adherence, measured as the number of detectable (>400) viral load measurements divided by the total number of viral load measurements after HAART initiation, was associated with prognosis in CHORUS (HR=0.20 comparing 82% and 3% viral loads suppressed (third and first quartiles), 95% CI=0.16–0.25). However, the model which included adherence did not account for the difference between predicted and observed ADE-free survival in the JHHCC (Table 2), and in fact made the

difference greater, as patients at CHORUS tended to have a higher percentage of detectable viral loads post-HAART (Table 1).

The largest discrepancy between the CHORUS and JHHCCs was the difference in lost to follow-up rates: 23% in CHORUS and only 5% in JHHCC. This difference appeared to be due in part to different definitions of lost to follow-up: 9 months without a visit for CHORUS vs. 18 months for JHHCC. If we used the 9-month lost to follow-up definition for JHHCC, then 22 individuals who died more than 9 months after their last visit were recorded as lost to follow-up. This slightly increased the ADE-free survival estimates for the JHHCC, but the ADE-free survival estimates still remained lower than the mean predicted ADE-free survival probabilities (Table 2).

It is possible that many of those lost to follow-up in the CHORUS cohort actually progressed to AIDS or died. However, based on a multivariable logistic regression model, those lost to follow-up in CHORUS tended to be male, white, and have higher aCD4 counts ($P < 0.05$ for all); characteristics associated with longer ADE-free survival. We were able to more closely examine those lost to follow-up while in care at the Comprehensive Care Center (CCC), Nashville, TN. CCC patients accounted for 38% (716/1891) of the CHORUS population. Of these patients, 13% (96/716) were lost to follow-up, which is lower than the 29% (344/1175) observed among non-CCC sites in CHORUS. Using a social security index death search, we found that 7 of our 96 lost to follow-up patients had died, but in all cases the date of death was greater than 9 months after the date of their last visit. Only two persons had dates of death less than 18 months after their last visit. We also discovered that 2 persons who had originally been censored at the study freeze date actually died 67 and 110 days before the close of the study. Therefore, from our further inspection of CCC data, we were only able to find two additional deaths. We were unable to obtain additional data on potential ADEs for those lost to follow-up.

A sensitivity analysis was performed to investigate the impact of under-ascertainment of ADE or death for those lost to follow-up. In order for the mean predicted ADE-free survival to be consistent (lie within the 95% confidence intervals) with the observed ADE-free survival at years 3, 4, and 5 using the corrected data (censoring the 22 JHHCC deaths as lost to follow-up and including the 2 additional CHORUS deaths), at least 50% of those lost to follow-up in CHORUS must have had an ADE or died within one year of their last visit (Table 2). In order to predict the same ADE-free survival as observed at 5 years, 90% of those lost to follow-up in CHORUS must have had an ADE or died within one year of their last visit.

4. Discussion

Using re-sampling techniques and an independent dataset from JHHCC, we have described the prognostic ability of a model to predict ADE-free survival based on patient characteristics at initiation of HAART. In short, our model was fair at ordering patients' risk of disease progression based on characteristics at HAART initiation, but was not good at predicting ADE-free survival more than 3 years after HAART initiation in an independent cohort with different characteristics.

Several measures of the model's predictive ability were remarkably consistent between those estimated internally using the bootstrap and those seen in the JHHCC data. Bootstrapping estimated that the proportion of all patient pairs in which the predictions and outcomes would be concordant would be approximately 63%; the concordance measure of our model in the JHHCC was 63%. Bootstrapping estimated that those patients predicted by the model to have low 3 year ADE-free survival probability would tend to do a little better than predicted, with

an estimated shrinkage coefficient of 0.60. This was also seen in the JHHCC data; the estimated shrinkage was 0.58.

However, the model was poor at predicting the ADE-free survival probability beyond 3 years after HAART initiation. It is expected that predictive accuracy deteriorates with time, and the JHHCC survival estimates 5 or more years after HAART initiation were based on a relatively small number of patients. Nonetheless, the difference between cohorts is still notable. Those in the JHHCC were much more likely to die than predicted by our model. It does not appear to be due to differing initial HAART regimens, subsequent adherence rates, or the high number of IDUs in the JHHCC. The JHHCC tended to have more advanced HIV disease at initiation of HAART, as measured by lower aCD4, lower %CD4, and higher HIV-1 RNA. Though these covariates were included in the prediction model, perhaps the CHORUS dataset from which the model was created did not include enough patients with similar characteristics, and thus did not create a good model for patients in the more advanced stages of HIV disease who initiate HAART, and/or for IDUs and non-whites. Although we believe standards of care were similar between sites, unmeasured differences could have influenced results. Between-site differences in the rate of disease progression could also be due to other unmeasured factors such as social deprivation and community-level mortality rates.

It seems probable that the discrepancy between predicted and observed ADE-free survival in the JHHCC was at least in part due to differences in lost-to-follow-up rates. Some of the discrepancy in rates of loss to follow-up were due to differences between the cohorts' standard definitions for loss to follow-up (9 and 18 months without a visit for CHORUS and JHHCC, respectively). Definitions of loss to follow-up vary substantially between cohorts (e.g., [24]), and this is an important source of heterogeneity that should be considered when taking results from one cohort and then applying them to another. However, when using a common definition for loss to follow-up (9 months without a visit) the rate of loss to follow-up was still substantially higher in CHORUS than JHHCC, and approximately 50% of those lost would have had to have had an ADE or died in order for the predicted ADE-free survival estimates (derived from CHORUS) to be consistent with the observed estimates in JHHCC.

As an aside, the mortality rate in the JHHCC was higher than seen elsewhere; the incidence of ADE-death over time in the CHORUS cohort was more consistent with previous reports [23-25].

There are other models for predicting ADE-free survival, most notably the model from the ART-CC [2,25]. Our models are similar in concept but constructed differently. ART-CC used parametric survival models (Weibull), we used semi-parametric models (Cox); ART-CC grouped continuous predictors into categories, we employed restricted cubic splines; ART-CC removed covariates that were poor predictors of progression from the model, we left them in; ART-CC chose a final prognostic model using a leave-one out cross-validation system, we used a model previously chosen based on Akaike's information criteria (no cross-validation) to evaluate the association between %CD4 and ADE/death. There are advantages and disadvantages to the different modeling approaches with regards to model flexibility, efficiency, parsimony, and clinical interpretation which have been discussed elsewhere [18-19,26-28]. We were unable to perform a head-to-head comparison of the models as one of the key predictors from ART-CC, clinical stage at HAART initiation, was not available. However, comparing the measures of prognostic accuracy for our model with those by the ART-CC [25], their model appeared to perform better. Their estimated c-statistic applied to independent data was 0.73 compared to ours of 0.63, perhaps reflecting the importance of including clinical disease stage; their model showed little evidence of over-fitting, whereas ours did, perhaps due to the overuse of spline terms and implying we should have fit a more parsimonious model; and the mean predicted ADE-free survival for the ART-CC model applied

to CASCADE data was quite close to the Kaplan-Meier estimates throughout the 3 years of follow-up. Importantly, it was for greater than 3 years of follow-up that our model had especially poor prediction when applied to JHHCC data.

Perhaps the most important information gleaned from this study was simply that HIV cohorts can differ greatly, and possibly in ways that are not measured by the several variables that are typically considered prognostic. This study highlighted some of the challenges of creating a model from one cohort and trying to predict observations in another. We believe that the difference between predicted and observed ADE-free survival >3 years after HAART initiation in the JHHCC was primarily due to differences between the CHORUS and JHHCC that no reasonable model would be able to overcome. These differences included substantially different loss-to-follow-up rates, patient demographics, and levels of immunosuppression at HAART initiation. As we continue to combine and apply information across cohorts, such as has been done in the ART-CC and is now being done in IeDEA, this cross-cohort heterogeneity needs to be recognized.

Acknowledgments

Financial support: Vanderbilt-Meharry Center for AIDS Research (NIH program 930 AI54999) and National Institutes of Health (grant K23 AT002508-01 to T.H., grant K24 A1065298 to T.R.S., and grants K24 DA00432 and R01 DA11602 to R.D.M.)

References

1. Mellors JW, Munoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med* 1997;126:946–954. [PubMed: 9182471]
2. Egger M, May M, Chene G, Phillips AN, Ledergerber B, Dabis F, et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet* 2002;360:119–129. [PubMed: 12126821]
3. Gange SJ, Kitahata MM, Saag MS, Bangsberg DR, Bosch RJ, Brooks JT, et al. Cohort profile: the North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD). *Int J Epidemiol* 2007;36:294–301. [PubMed: 17213214]
4. McGowan CC, Cahn P, Gotuzzo E, Padgett D, Pape JW, Wolff M, et al. Cohort profile: Caribbean, Central and South America Network for HIV research (CCASAnet) collaboration within the International Epidemiologic Databases to Evaluate AIDS (IeDEA) programme. *Int J Epidemiol* 2007;36:969–976.
5. Sterling TR, Chaisson RE, Moore RD. Initiation of highly active antiretroviral therapy at CD4+ T lymphocyte counts of >350 cells/mm³: disease progression, treatment durability, and drug toxicity. *Clin Infect Dis* 2003;36:812–5. [PubMed: 12627368]
6. Hogg RS, Yip B, Chan KJ, et al. Rates of disease progression by baseline CD4 cell count and viral load after initiating triple-drug therapy. *JAMA* 2001;286:2568–77. [PubMed: 11722271]
7. Phillips AN, Staszewski S, Weber R, et al. HIV viral load response to antiretroviral therapy according to the baseline CD4 cell count and viral load. *JAMA* 2001;286:2560–7. [PubMed: 11722270]
8. Sterling TR, Chaisson RE, Keruly J, Moore RD. Improved outcomes with earlier initiation of highly active antiretroviral therapy among human immunodeficiency virus-infected patients who achieve durable virologic suppression: longer follow-up of an observational cohort study. *J Infect Dis* 2003;188:1659–65. [PubMed: 14639536]
9. Palella FJ Jr, Deloria-Knoll M, Chmiel JS, et al. Survival benefit of initiating antiretroviral therapy in HIV-infected persons in different CD4+ cell strata. *Ann Intern Med* 2003;138:620–6. [PubMed: 12693883]
10. Opravil M, Ledergerber B, Furrer H, et al. Clinical efficacy of early initiation of HAART in patients with asymptomatic HIV infection and CD4 cell count > 350 × 10⁶ /l. *AIDS* 2002;16:1371–81. [PubMed: 12131214]

11. The Antiretroviral Therapy Cohort Collaboration. Rates of disease progression according to initial highly active antiretroviral therapy regimen: a collaborative analysis of 12 prospective cohort studies. *J Infect Dis* 2006;194:612–622. [PubMed: 16897660]
12. Hulgán T, Shepherd BE, Raffanti SP, Fusco JS, Beckerman R, Barkanic G, Sterling TR. Absolute count and percentage of CD4+ lymphocytes are independent predictors of disease progression in HIV-infected person initiating highly active antiretroviral therapy. *Journal of Infectious Diseases* 2007;195:425–431.
13. Becker SL, Raffanti SR, Hansen NI, et al. Zidovudine and stavudine sequencing in HIV treatment planning: findings from the CHORUS HIV cohort. *J Acquir Immune Defic Syndr* 2001;26:72–81. [PubMed: 11176271]
14. Moore RD. Understanding the clinical and economic outcomes of HIV therapy: the Johns Hopkins HIV clinical practice cohort. *J Acquir Immune Defic Syndr Hum Retrovirol* 1998;17(Suppl 1):S38–41. [PubMed: 9586651]
15. Centers for Disease Control. Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Recomm Rep* 1993;1992;41(RR17):1–19.
16. Fusco, GP.; Justice, AC.; Becker, SL., et al. Strategies for obtaining consistent diagnoses for cause of death in an HIV observational cohort study [#5584].. Presented at: 2002 XIV International AIDS Conference; Barcelona, Spain.
17. Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994;81:515–526.
18. Harrell, FE. *Regression Modeling Strategies, with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer; New York: 2001.
19. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996;15:361–387. [PubMed: 8668867]
20. Shepherd BE. The cost of checking proportional hazards. *Statistics in Medicine* 2008;27:1248–1260. [PubMed: 17708517]
21. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association* 1983;78:316–331.
22. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*. New York; Chapman & Hall: 1993.
23. Sterne JA, May MT, Sabin CA, Phillips AN, Costagliola D, Chene G, Justice AC, et al. Importance of baseline prognostic factors with increasing time since initiation of highly active antiretroviral therapy: collaborative analysis of cohorts of HIV-1 infected patients. *J Acquir Immune Defic Syndr*. in press
24. Rosen S, Fox MP, Gill CJ. Patient retention in antiretroviral therapy programs in Sub-Saharan Africa: a systematic review. *PLoS Med* 2007;4(10):e298. [PubMed: 17941716]
25. May M, Porter K, Sterne JA, Royston P, Egger M. Prognostic model for HIV-1 disease progression in patients starting antiretroviral therapy was validated using independent data. *Journal of Clinical Epidemiology* 2005;58:1033–1041. [PubMed: 16168349]
26. May M, Royston P, Egger M, Justice AC, Sterne JA, ART Cohort Collaboration. Development and validation of a prognostic model for survival time data: application to prognosis of HIV positive patients treated with antiretroviral therapy. *Statistics in Medicine* 2003;23:2375–2398. [PubMed: 15273954]
27. Altman DG, Royston P. The cost of dichotomizing continuous variables. *BMJ* 2006;332:1080. [PubMed: 16675816]
28. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* 2006;25:127–141. [PubMed: 16217841]

Figure 1.

A. Predicted vs. observed 3-year ADE-free survival in the CHORUS cohort. Patients were divided into 5 risk categories based on predicted ADE-free survival. Predicted values are mean predicted 3-year ADE-free survival within each of the five risk categories and observed values are Kaplan-Meier estimates and 95% confidence intervals for the same five groups of HAART initiators. X represents what we expect to observe in independent data, based on our internal validation procedure using the bootstrap. The dotted line represents perfect predictive accuracy.

B. Predicted vs. observed 3 year ADE-free survival in the JHHCC.

CHORUS	1891	1640	1404	1218	1044	854
JHHCC	716	570	448	326	232	158
	Number at Risk					

Figure 2. Mean predicted vs. observed ADE-free survival probability in JHHCC and CHORUS. Dotted lines represent 95% confidence intervals of Kaplan-Meier estimates.

Figure 3.
Rates of ADE/death over time for IDUs and non-IDUs in CHORUS and JHHCC.

Table 1

Characteristics of persons at the start of HAART in CHORUS and JHHCC

	CHORUS (n=1891)	JHHCC (n=716)
Baseline %CD4 *	16 (9, 24) **	12 (4, 19)
Baseline aCD4 *	240 (107, 384)	133 (28.8, 271)
Age (years) ***	38 (33, 43)	39 (34, 44)
Male *	1692 (90)	467 (65)
White Race *	1364 (72)	128 (18)
Baseline Log-Viral Load *	4.7 (4.3, 5.2)	4.9 (4.4, 5.4)
Prior ART *	1001 (53)	289 (40)
IDU *	108 (5.7), NA=115 (6%)	327 (46), NA=40 (5.6%)
Follow-up years *	4.6 (1.9, 6.9)	2.7 (1.2, 4.8)
ADE/death *	468 (24.7)	270 (38)
ADE	281 (15)	84 (12)
death	187 (10)	186 (26)
AIDS related	98 (5.2)	99 (13.8)
not AIDS related	41 (2.2)	38 (5.3)
unknown cause	48 (2.5)	49 (6.8)
Censored	1423 (75.3)	716 (62)
lost to follow-up	440 (23)	35 (5)
no event by Aug 1, 2005	983 (52)	411 (57)
First HAART Regimen *		
PI	1281 (68)	362 (54)
NNRTI	404 (21)	271 (40.5)
PI/NNRTI	31 (1.6)	0 (0)
NRTI	203 (10.7)	36 (5.5)
Unknown	0 (0)	47 (6.6)
Proportion Suppressed Viral Loads post-HAART initiation *	0.38 (0.00, 0.77), NA=69	0.50 (0.08, 0.90)

NA = Not available.

* p<0.0001 comparing CHORUS with JHHCC using Wilcoxon rank sum test for continuous variables, Pearson's chi-square test for categorical.

** Median (Interquartile range) reported for continuous variables; N (%) for categorical.

*** p=0.09

Table 2

	3 years	4 years	5 years
Observed JHU ADE-free survival (95% CI)	.73 (.69, .77)	.64 (.60, .68)	.57 (.52, .62)
Observed JHU ADE-free survival with lost to follow-up=9 mo without visit (95% CI)	.75 (.71, .78)	.67 (.62, .71)	.60 (.55, .64)
Mean Predicted JHU ADE-free survival			
Original Model	.76	.72	.69
Model including first HAART Regimen	.76	.72	.69
Model stratifying by IDU	.78	.73	.69
Model including proportion of undetectable viral loads	.84	.81	.78
Model assuming 50% of CHORUS patients lost to follow-up had event within year	.74	.68	.64
Model assuming 90% of CHORUS patients lost to follow-up had event within year	.72	.65	.60