

Dynamic Predictions From Longitudinal CD4 Count Measures And Time To Death of HIV/AIDS Patients Using a Bayesian Joint Model

Feysal Kemal Muhammed (✉ feysal.kemal@gmail.com)

Hawasa University

Aboma Temesgen Sebu

Haramaya University

Anne M Presanis

University of Cambridge

Denekew Bitew Belay

Bahir Dar University

Research Article

Keywords: Joint model, Dynamic predictions, Longitudinal, Time-to-event, and Bayesian Model Averaging

Posted Date: September 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-873385/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Personalised or stratified medicine has played an increasingly important role in improving bio-medical care in recent years. A Bayesian joint modelling approach to dynamic prediction of HIV progression and mortality allows such individualised predictions to be made for HIV patients, based on monitoring of their CD4 counts. This study aims to provide predictions of patient-specific trajectories of HIV disease progression and survival.

Methods: Longitudinal data on 254 HIV/AIDS patients who received ART between 2009 and 2014, and who had at least one CD4 count observed, were employed in a Bayesian joint model of disease progression, as measured by CD4 counts, and survival, to obtain individualised dynamic predictions of both processes that were updated at each visit time in the follow-up period. Different forms of association structure that relate the longitudinal CD4 biomarker and time to death were assessed; and predictions were averaged over the different models using Bayesian model averaging.

Results: A total of 254 subjects were observed in the dataset with a median age of 30 years (interquartile range, IQR, 26–38). The individual follow-up times ranged from 1 to 120 months, with a median of 22 months and IQR 7–39 months. The median baseline CD4 count was 129 cells/mm³ (IQR 61–247 cells/mm³). From the joint model with highest posterior weight, subjects whose functional status was working were significantly associated with a higher baseline CD4 count ($\beta = 1.86$; 95% CI: 0.65–3.04) whereas subjects who were bedridden were significantly associated with a lower baseline CD4 count (estimated effect $\beta = -3.54$; 95% CI: -5.65, -1.39), compared to ambulatory patients. A unit increase in weight of the individual increased the mean square root CD4 measurement by 0.06. The estimates of the association structure parameters from all three models considered indicated that the HIV mortality hazard at any time point is associated with the current underlying value of the CD4 count at the same time point. The model with highest posterior weight also had a time-dependent slope, indicating that HIV mortality is also associated with the rate of change in CD4 count. From both the model-averaged predictions and the highest posterior weight model alone, an increase in CD4 count was predicted at different visit times from the dynamic predictions. It was also found that there was an increase in the width of prediction intervals as time progressed.

Conclusions: Functional status, weight and alcohol intake are important contributing factors that affect the mean square root of CD4 measurements. For this particular dataset, model averaging the dynamic predictions resulted in only one of the hypothesised association structures having non-zero weight at the majority of time points for each individual. The predictions were therefore similar whether we averaged them over models or derived them from the highest posterior weight model alone. We also observed that the parameter estimates in the both the CD4 count and survival sub-models showed slight variability between the postulated association structures.

Background

Ethiopia has been severely affected by the HIV epidemic for the last three decades. Estimates of HIV prevalence in the general population remain low despite an increasing number of people living with HIV/AIDS and taking antiretroviral therapy (ART). The 2016 Ethiopian Demographic and Health Survey reports that the prevalence of adult HIV in Ethiopia is 0.9%, but with major differences by region (4.8 % in Gambella, 3.4 % in Addis Ababa, and 0.4 % in SNNPR), and by type of region (2.9 % urban versus 0.4 % rural). According to the latest UNAIDS Spectrum-derived estimates, 613,803 people were living with HIV (PLHIV) in Ethiopia by the end of 2017, of whom 443,213 (72%) were on treatment [1]. There has been an increasing interest in personalized medical research in recent years, including in the HIV field. Monitoring of a biomarker of disease progression can allow doctors to customize treatment decisions to the characteristics of the patient, to improve medical care and survival[2]. For example, monitoring a HIV patient's CD4 counts allows a clinician to decide when to start and adjust anti-retroviral treatment[3]. However, analysing a longitudinal biomarker measuring disease progression, such as CD4 count, and a survival outcome separately can lead to biased estimates of both the biomarker and survival processes, as such an analysis ignores the dependence between the repeated longitudinal measurements and the survival process. Joint modelling of longitudinal and survival data is preferred to separate analyses both to optimally use the available information and to obtain unbiased estimates of parameters describing both processes [4, 5]. Joint models are versatile instruments in deriving probabilities of survival and forecasts for future levels of biomarkers [6]. Temesgen et al [7] modelled jointly the weight of TB patients and CD4 counts which act as biomarker for HIV disease progression. A Bayesian approach to joint modelling can be used to derive dynamic individualised predictions of disease progression and survival [8]. Barrett and Su [9] proposed a new flexible joint model with subject-specific penalized splines (P-splines) used to characterize the co-evolution of the longitudinal and time-to-event processes. Rizopoulos et al [10] used Bayesian Model Averaging (BMA) for combining dynamic predictions from multiple alternative joint models for longitudinal and time-to-event data. The aim of this study is to derive individual-level predictions of both CD4 count progression and HIV survival, based on a collection of possible models simultaneously, combining them using Bayesian model averaging.

Methods

2.1 Data, Setting and Participants

This study used data from HIV patients from Jimma University Specialized Hospital, Ethiopia, in order to predict the probabilities of HIV-related survival over time by jointly modelling the longitudinal CD4 count representing HIV disease progression and time-to-event processes. While the population comprised a total of 854 patients who first received ART between 2009 and 2014, the specific HIV data for this analysis came from a longitudinal study of the subset of 254 patients aged at least 18 years who had at least one measurement of CD4 count. These patients were followed up for a maximum of 10 years, with visit times at which CD4 counts were recorded occurring approximately every 6 months. A histogram of observed CD4 counts in Figure A1 (Appendix III) shows the data fail to fulfil a normality assumption: we therefore

use a square root transformation of the observed CD4 counts to achieve normality (Figure A2, Appendix III).

2.2. Outcome

The two outcome variables considered for this study were the survival outcome, i.e., time from attendance date to death, measured in months, and the observed CD4 counts, measured in cells/mm³ of blood. WHO's clinical stage of disease, the functional status of patients (ambulatory, bedridden or working), weight, smoking status, alcohol consumption status, drug use status and marital status were the covariates considered. Further details about the data are found in Temesgen et al [7].

2.3 Statistical analysis

Since HIV survival is known to be dependent on disease progression, such as measured by CD4 count [3, 7, 11], careful consideration of the statistical method used to relate the two outcome variables is important. A key characteristic of HIV disease progression is its dynamic nature: the rate of progression is not only different from patient to patient, but also dynamically changes in time for the same patient. Thus, the true potential of the CD4 count biomarker in describing disease progression and its association with survival can only be exploited when repeated measurements of CD4 count are considered in the analysis. The structure of the dependence between the two outcomes is not fully known, and may vary between populations, and even over time within a single patient. To address research questions involving characterisation of the association structures between repeated measures and event times, a class of statistical models has been developed known as joint models for longitudinal and time-to-event data [2, 12–14]. Briefly, a mixed effects model is proposed for the longitudinal biomarker observations, of a general form $y_i(t) = m_i(t) + \varepsilon_i(t)$ where the mean $m_i(t)$ is the (linear) predictor comprising both fixed and random effects, and $\varepsilon_i(t)$ is a normally distributed error term. Simultaneously, a standard survival model is posited for the time-to-event data, regressed both on covariates and on the mean $m_i(t)$ of the longitudinal process. Different forms of the regression on $m_i(t)$ are possible, including regressing only on the current value of the mean, regressing on both the current value and rate of change, or instead regressing on the random effects that are included in $m_i(t)$, for example. Given the random effects in $m_i(t)$, both the longitudinal and survival processes are assumed independent, as are the longitudinal responses of each individual. The random effects therefore account both for the association between the longitudinal and the survival outcomes and the correlation between the repeated measurements in the longitudinal process.

Extensions of joint models such as dynamic predictions and accuracy measures have also been implemented [10, 15, 16]. Dynamic prediction is a method for updating predictions ahead in time of both the longitudinal and survival processes, whenever a new measurement of the longitudinal biomarker is taken. Here three joint models of the evolution of the CD4 count process and HIV survival are fitted to the data in a Bayesian framework, each with a different association structure. Dynamic predictions are derived from each of the three models and are combined using Bayesian model averaging [10, 17]. The Bayesian approach to joint modelling [18, 19] was implemented using the JMbayes package in R version

1.2.5033. Full details on the formulation of these joint models and dynamic prediction can be found in Appendix I, but are briefly summarised below.

To capture the non-linear subject-specific evolution of CD4 counts (Figs. 1; 2), a flexible specification of the subject-specific square-root CD4 trajectories was adopted, using natural cubic splines of time. Specifically, the linear mixed model took the form:

$$y_i(t) = m_i(t) + \varepsilon_i(t) = \beta_0 + \beta_1 \text{FNS}_i + \beta_2 \text{Alcohol}_i + \beta_3 \text{MS}_i + \beta_4 \text{CS}_i + \beta_5 \text{wt}_i + \beta_6 B_1(t) + \beta_7 B_2(t) + \beta_8 B_3(t) + b_{i0} + b_{i1} B_1(t) + b_{i2} B_2(t) + b_{i3} B_3(t) + \varepsilon_i(t)$$

where $B_n(t)$ denotes the B-spline basis for a natural cubic spline. FNS, Alcohol, MS, CS, and wt are the variables functional status, alcohol intake, marital status, clinical stage and weight respectively. The β 's are fixed effects whereas the b's are random effects. For the survival process we consider three relative risk models, each positing a different association structure between the two processes, namely:

$$M_1(t):h_1(t) = h_0(t)\exp\{\gamma_1 \text{FNS}_i + \gamma_2 \text{Alcohol}_i + \gamma_2 \text{MS}_i + \gamma_3 \text{CS}_i + \gamma_4 \text{wt}_i + \alpha_1 m_i(t)\},$$

$$M_2(t):h_2(t) = h_0(t)\exp\{\gamma_1 \text{FNS}_i + \gamma_2 \text{Alcohol}_i + \gamma_2 \text{MS}_i + \gamma_3 \text{CS}_i + \gamma_4 \text{wt}_i + \alpha_1 m_i(t) + \alpha_2 m_i(t)\},$$

$$M_3(t):h_3(t) = h_0(t)\exp\{\gamma_1 \text{FNS}_i + \gamma_2 \text{Alcohol}_i + \gamma_2 \text{MS}_i + \gamma_3 \text{CS}_i + \gamma_4 \text{wt}_i + \alpha_1 b_{i0} + \alpha_2 b_{i1} + \alpha_3 b_{i2} + \alpha_4 b_{i3}\}$$

where the baseline hazard $h_0(t)$ is approximated with splines (see Eq. (3) of Appendix I), $m_i(t)$ is the current true value of the CD4 count trajectory, $m_i(t)$ is the slope of the trajectories at time t (rate of change in CD4 count), γ are regression parameters of the survival model and α are parameters describing the strength of the association between the CD4 count and survival processes.

2.4 Individualized dynamic predictions

Based on each joint model, prediction of survival probabilities and future CD4 counts for a new individual j who has a set of longitudinal square-root CD4 counts $Y_j(t) = \{y_{jl}(s); 0 \leq t_{jl} \leq t, l=1, \dots, n_j\}$ and a vector of baseline covariates w_j is required. For any time $u > t$, the focus of interest is in predicting both the conditional probability $\pi_j(u|t)$ that subject j will survive at least up to u and his/her predicted CD4 count at u . At each time of interest (e.g. a clinic visit time) $t' < t < u$, these predictions are dynamically updated, as extra information is recorded for the patient. That is, the prediction $\omega_j(u|t)$ of the square-root CD4 count $y_j(u)$ that is based on the information available up to time t , can be updated at time t' to produce a new prediction $\omega_j(u|t')$ that uses the additional longitudinal information up to the latter time point t' . Under the Bayesian joint modelling framework, both predictions $\pi_j(u|t)$ and $\omega_j(u|t)$ are based on the posterior predictive distribution, as given in Appendix I.

Standard model selection techniques for choosing between the three association structures may prefer different models, depending which model selection criteria is used[20, 21], particularly in contexts where different association structures may produce better predictions for different individuals at different time

points. BMA [10, 16] explicitly takes into account model uncertainty by applying Bayesian inference to model selection. Each model is given a prior weight, in this case assuming each association structure is equally likely, and the resulting posterior model weights are used to average over the estimates. Here, following [10], instead of averaging estimates over the association structures, the predictions $\pi_j(u|t)$ and $\omega_j(u|t)$ are averaged over the different association structures. This BMA approach can produce less risky predictions via a straightforward model choice criteria[22].

Results

Figures 1 and 2 show that some individual longitudinal CD4 count trajectories had strong nonlinear patterns, motivating the flexible structure of the joint models proposed in Section 2.

A total of 254 subjects were observed in the dataset with a median age of 30 years (interquartile range, IQR, 26–38). The individual follow-up times ranged from 1 to 120 months, with a median of 22 months and IQR 7–39 months. The median baseline CD4 count of the subjects included in the analysis was 129 cells/mm³ (IQR 61–247 cells/mm³). The square-root transformed CD4 count of the subjects ranged from 2.5 to 27.4 with a first quartile of 7.8, a median of 11.4 and a third quartile of 15.7.

Table A1 (Appendix II) summarises the demographics of the study population. Based on WHO's clinical stage categories of HIV progression, the majority were at an advanced stage of disease (47% at Stage III, 36% at Stage IV). With regard to the functional status of subjects, 46% were able to work, 47% were ambulatory, and 6% were bedridden. Of the total deaths, 47% occurred in ambulatory patients, while 24% occurred in working patients. 46% of patients were female, 54% were male. Around half of the subjects (45%) had received primary education, while 5% had achieved tertiary education. 43% of individuals were married, while 7% were divorced.

Table A2 (Appendix II) shows estimates and their corresponding 95% confidence intervals for the parameters in the longitudinal and survival sub-models for each of the three different association structures. In the models including current CD4 count in the association structure (Models 1 and 2), the association parameters indicate that HIV survival at any time t is associated with the current underlying CD4 count at the same time point. The parameter estimates in the relative risk models and in the linear mixed models show slight variability between the posited association structures. Model 2, where the association is expressed in terms of both the current value of the CD4 count process and the rate of change in CD4 count, has the lowest Deviance Information Criterion. Furthermore, in the model-averaged predictions (Table A3, Appendix II for selected individuals), Model 2 is the only model with non-zero weight for the vast majority of time points for each individual, with only a few specific time points where any other model has any weight. Convergence of the Markov chain Monte Carlo algorithm for this selected Model 2 is demonstrated in Appendix III (Figure A3). The Model 2 regression coefficient estimates for the CD4 count process suggest that subjects whose functional status was working were significantly associated with a higher baseline CD4 count ($\beta = 1.86$; 95% CI: 0.65 3.04) whereas those whose functional status was bedridden were significantly associated with lower baseline CD4 count ($\beta =$

-3.54; 95% CI: -5.65 -1.39), compared with the baseline ambulatory functional status group. A unit increase in weight of the patient increased the mean square-root CD4 measurements by 0.06 and linear time, quadratic time and cubic time had positive effects on the mean rate of change of the square-root CD4 measurements. Functional status, alcohol consumption and weight also had significant effects on hazard of HIV mortality: 0.42 times lower for patients able to work compared to those with ambulatory functional status; 1.63 times higher for patients who consume alcohol compared to the non-alcohol drinking group at baseline; and 0.98 times lower for each unit increase in weight. A unit increment in the patient-specific current square-root CD4 count decreases the hazard of HIV mortality 0.93 times ($\beta = -0.07$; 95% CI: -0.13 -0.005). Despite Model 2 having lower DIC than Model 1, where the only difference is the inclusion of the rate of change in CD4 count in the association structure, the estimated association with the rate of change is marginally non-significant ($\beta = -2.56$; 95% CI: -4.53 0.31). However, the point estimate would imply that for a unit increase in the rate of change (linear slope) in square-root CD4 count, the hazard of HIV mortality is decreased 0.08 times.

Table A3 (Appendix II) shows model-averaged predictions for five individuals, who have been selected randomly from the group of patients who were still alive at the end of the study. The table also shows the time-dependent subject-specific BMA weights for the three joint models. The dynamic predictions are updated at each visit time for each patient, using the CD4 counts observed up to the current visit time to predict both future CD4 count trajectories and HIV survival. As noted, only Model 2 (a model with extended association structure) has non-zero weight at the majority of time points, resulting in model-averaged dynamic predictions that are similar to the Model 2 predictions. Figures 3 and 4 display these predictions for one example patient (id 2151). This patient's observed and expected CD4 counts increase over the study period.

The vertical dotted lines in both Figures 3 and 4 represent the time point of the last square-root CD4 count observation. The left sides of the panels present the observed square root of CD4 counts and the right shows the estimated individual longitudinal trajectories of CD4 count for randomly selected subject 2151. The dashed lines represent the corresponding 95% point wise confidence intervals. The dynamic survival probabilities and dynamic predictions of CD4 counts for the other four randomly selected patients are given in Appendix III.

Discussion

We have fitted three joint models incorporating different association structures that relate HIV survival to the longitudinal trajectories of patient CD4 counts; then used Bayesian model averaging techniques to select the model on which to base dynamic predictions. For all patients in our study, Model 2, which considers both the current estimated value of the CD4 count trajectory and its rate of change, was the only model with non-zero weight at the majority of time points, with only a few time points for a few individuals where other models had non-zero weight. The model averaging therefore effectively selected only Model 2 for prediction. Fitting Model 2 alone, without model averaging, resulted in similar predictions, as expected given the near-zero weights for Models 1 and 3.

The estimates of the association parameters in Model 2, together with the dynamic survival probability plots, showed that as CD4 count increases, the rate of decrease in the survivor function slows, indicating that time to death is longer for those with higher CD4 count and for those with a positive rate of change in CD4 counts. Note that we found an increase in the width of the prediction intervals for the future CD4 counts as time progressed. Note also that Rizopoulos et al. (2017) [16] state that differences in prediction performance between different specifications of the association structure might be due to the dependence structure not being correctly specified. Rizopoulos et al (2014) [10] compared dynamic prediction results from five joint models with different parameterizations in the survival sub-model and found that the predicted conditional survival probabilities showed considerable variability between the six parameterizations. As a response to the challenge of different results from different parameterisations, Barrett & Su use a different approach, using a flexible non- or semi-parametric specification of the association structure, rather than a parametric specification, to avoid mis-specifying the dependence structure. Although BMA has some interesting features, it has some potential drawbacks as well, including being computationally intensive, since it requires fitting all the models we wish to average. Other approaches to individualised dynamic prediction include stacking and pseudo-BMA [23]; and landmarking [24]. A number of other papers have also compared prediction models for dynamic prediction [25]; [26]; [27].

Conclusions

The main aim of this study was to see how dynamic predictions from joint models incorporating different association structures can be combined using Bayesian model averaging. The result from this study showed that there is some variability between the three parameterizations we used in terms of parameter estimates. We also demonstrated that a reasonably accurate specification of the CD4 count biomarker trajectory was essential to provide good predictions with joint models, with only one model having high weight in the averaged predictions. Many studies used standard likelihood information criteria to decide on the selection of appropriate joint model and make predictions. Although, in the case of our specific HIV dataset, we found that Bayesian model averaging gave similar dynamic predictions to those obtained from a single model (Model 2) alone, we recognise that in other applications, a single prognostic model may not be adequate, in which case BMA may provide a good solution to sensitivity of predictions to model specification, accounting for model uncertainty. Investigation of different parameterisations and sensitivity of parameter estimates and predictions to model specification is essential to the provision of individual-level predictions of patient progression and survival.

Abbreviations

HIV: Human Immunodeficiency Virus; CD4: cluster of differentiation 4; AIDS: Acquired Immune Deficiency Syndrome; ART: Antiretroviral therapy; IQR: Interquartile range; EDHS: Ethiopian Demographic and Health Survey; SNNPR: Southern Nations, Nationalities, and Peoples' Region; UNAIDS: United Nations Acquired Immune Deficiency Syndrome; PLHIV: People living with HIV; TB: Tuberculosis; BMA: Bayesian Model

Averaging ; WHO: World Health Organization; MCMC: Markov chain Monte Carlo; AIC: Akaike information criterion; BIC: Bayesian information criterion .

Declarations

Acknowledgments

We would like to express our profound gratitude to the management of Jimma University Specialized Hospital for allowing us to use the medical register from which we extracted the data used in this study. We would also like to thank Dr. Jessica Barrett for providing her guidance and advice on the methods used.

Authors' contributions

ATS collected the data, FKM designed the study, analysed the data and wrote the manuscript. AMP and DBB advised on analysis and edited the manuscript. All authors read and approved the final manuscript.

Authors' information

FKM is a Ph.D student; AMP is a Senior Investigator Statistician at the MRC Biostatistics Unit, University of Cambridge, United Kingdom, DBB is an assistant professor at Department of Statistics, Bahir Dar University, and ATS is an assistant professor at Department of Statistics, Haramaya University, Dire Dawa , Ethiopia.

Availability of data and materials

The datasets analysed in this study are not publicly available due to patient confidentiality, but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Secondary data which didn't contain personal identifiers were used for our study. Th secondary data were collected from patient cards anonymously with generated identifier codes. Ethical approval for the study was obtained from Department of Statistics of Jimma University. The department ethics committee approved this procedure.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest

Funding

No funding was obtained for this study.

References

1. ICF, C.S.A.-C.E.a., *Ethiopia Demographic and Health Survey 2016*. 2017.
2. Rizopoulos, D., *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R (1st ed.)*. 2012.
3. Dessie, Z.G., et al., *Modelling of viral load dynamics and CD4 cell count progression in an antiretroviral naive cohort: using a joint linear mixed and multistate Markov model*. BMC Infect Dis, 2020. **20**(1): p. 246.
4. Asar, Ö., et al., *Joint modelling of repeated measurement and time-to-event data: an introductory tutorial*. International Journal of Epidemiology, 2015. **44**(1): p. 334–344.
5. Rizopoulos, D. and P. Ghosh, *A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event*. Statistics in Medicine, 2011. **30**(12): p. 1366–1380.
6. Long, J.D. and J.A. Mills, *Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease*. BMC Medical Research Methodology, 2018. **18**(1): p. 138.
7. Temesgen, A., A. Gurmesa, and Y. Getchew, *Joint Modeling of Longitudinal CD4 Count and Time-to-Death of HIV/TB Co-infected Patients: A Case of Jimma University Specialized Hospital*. Annals of Data Science, 2018. **5**(4): p. 659–678.
8. Rizopoulos, D., *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. 2012: p. 275.
9. Barrett, J. and L. Su, *Dynamic predictions using flexible joint models of longitudinal and time-to-event data*. Stat Med, 2017. **36**(9): p. 1447–1460.
10. Rizopoulos, D., et al., *Combining Dynamic Predictions From Joint Models for Longitudinal and Time-to-Event Data Using Bayesian Model Averaging*. Journal of the American Statistical Association, 2014. **109**(508): p. 1385–1397.
11. Tsiatis, A.A.a.D., M., *Joint Modeling of Longitudinal and Time-to-Event Data: An Overview*. Statistica Sinica, 2004. **14**: p. 809–834.
12. Henderson, R., P. Diggle, and A. Dobson, *Joint modelling of longitudinal measurements and event time data*. Biostatistics, 2000. **1**(4): p. 465–80.
13. Hsieh, F., Y.-K. Tseng, and J.-L. Wang, *Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited*. 2006. **62**(4): p. 1037–1043.
14. Wulfsohn, M.S. and A.A. Tsiatis, *A joint model for survival and longitudinal data measured with error*. Biometrics, 1997. **53**(1): p. 330–9.
15. Barrett, J. and L. Su, *Dynamic predictions using flexible joint models of longitudinal and time-to-event data*. 2017. **36**(9): p. 1447–1460.

16. Rizopoulos, D., G. Molenberghs, and E. Lesaffre, *Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking*. *Biom J*, 2017. **59**(6): p. 1261–1276.
17. Hoeting, J.A., et al., *Bayesian Model Averaging: A Tutorial*. *Statistical Science*, 1999. **14**(4): p. 382–401.
18. Brown, E., J. Ibrahim, and V.J.B. DeGruttola, *A flexible B-spline model for multiple longitudinal biomarkers and survival*. 2005. **61** **1**: p. 64–73.
19. Ibrahim, J.G., M.-H. Chen, and D. Sinha, *Bayesian Survival Analysis*, in *Encyclopedia of Biostatistics*. 2005.
20. Akaike, H., *A new look at the statistical model identification*. *IEEE Transactions on Automatic Control*, 1974. **19**(6): p. 716–723.
21. Schwarz, G., *Estimating the Dimension of a Model*. *The Annals of Statistics*, 1978. **6**(2): p. 461–464.
22. Fragoso, T.M., W. Bertoli, and F. Louzada, *Bayesian Model Averaging: A Systematic Review and Conceptual Classification*. *International Statistical Review*, 2018. **86**(1): p. 1–28.
23. Barigou, K., et al. *Bayesian model averaging for mortality forecasting using leave-future-out validation*. 2021.
24. Ferrer, L., H. Putter, and C. Proust-Lima, *Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment*. *Statistical Methods in Medical Research*, 2019. **28**(12): p. 3649–3666.
25. Goldstein, B.A., et al., *A comparison of risk prediction methods using repeated observations: an application to electronic health records for hemodialysis*. *Statistics in Medicine*, 2017. **36**(17): p. 2750–2763.
26. Huang, X., et al., *A two-stage approach for dynamic prediction of time-to-event distributions*. *Stat Med*, 2016. **35**(13): p. 2167–82.
27. Sweeting, M.J., *Using predictions from a joint model for longitudinal and survival data to inform the optimal time of intervention in an abdominal aortic aneurysm screening programme*. *Biometrical Journal*, 2017. **59**(6): p. 1247–1260.
28. Cox, D.R., *Regression Models and Life-Tables*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1972. **34**(2): p. 187–220.
29. Laird, N.M. and J.H. Ware, *Random-Effects Models for Longitudinal Data*. *Biometrics*, 1982. **38**(4): p. 963–974.
30. Harville, D.A., *Matrix Algebra From a Statistician's Perspective*. *Technometrics*, 1998. **40**(2): p. 164–164.
31. Verbeke, G.a.G.M., *Linear Mixed Models For Longitudinal Data*. 2005.
32. Hsieh, F., T. Yi-Kuan, and J.-L. Wang, *Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited*. *Biometrics*, 2006. **62**(4): p. 1037–1043.
33. Lang, S. and A. Brezger, *Bayesian P-Splines*. *Journal of Computational and Graphical Statistics*, 2004. **13**(1): p. 183–212.

Figures

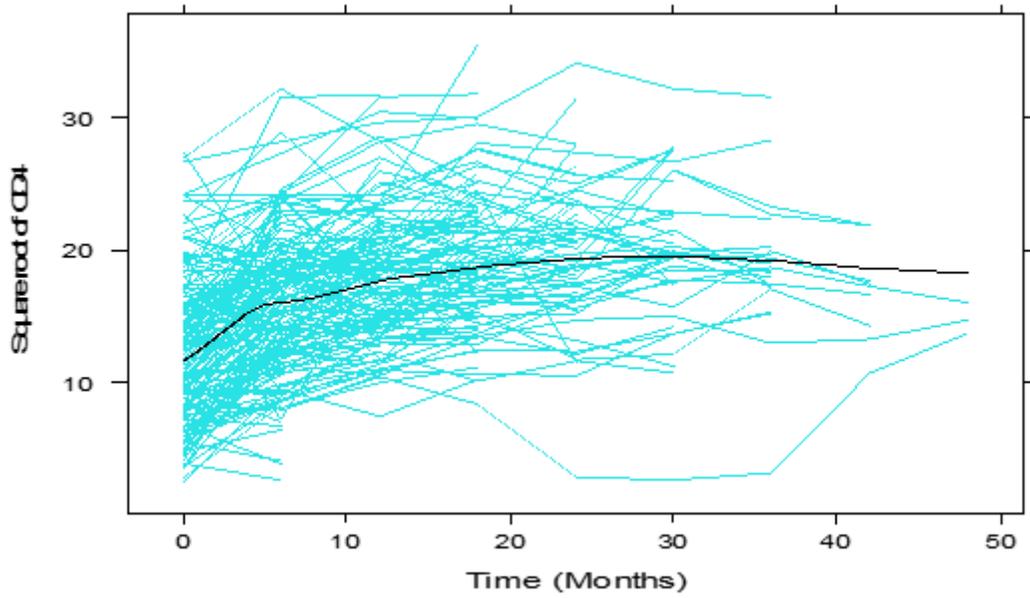


Figure 1

Individual and mean profiles of observed CD4 count data over time.

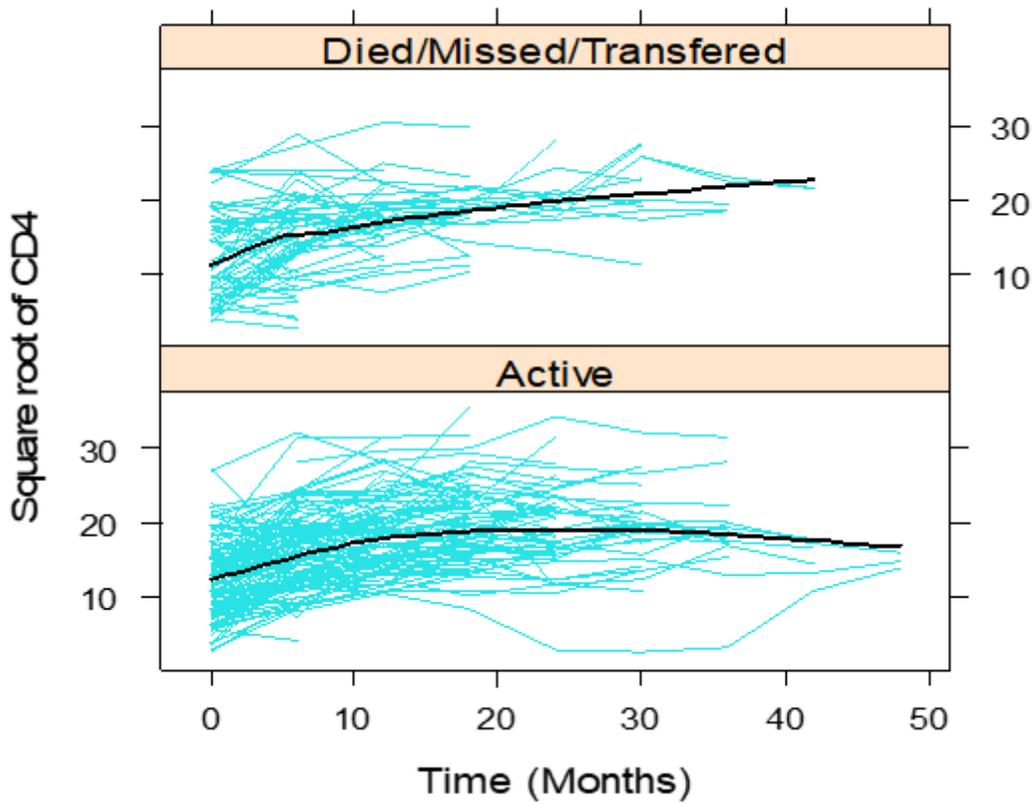


Figure 2

Individual and mean profiles of square root CD4 counts over time, categorized by status

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix.docx](#)