
saMRC
advancinglife


UHASSELT
I-BioStat
KU LEUVEN
Interuniversity Institute for Biostatistics
and statistical Bioinformatics

A novel approach to estimation of the time to biomarker threshold: Applications to HIV

Pharmaceutical Statistics, Volume 15, Issue 6, Pages 541-549, November/December 2016
PSI Journal Club – 22 March 2017

Tarylee Reddy [1,2] , **Geert Molenberghs** [2,3], **Edmund Njeru Njagi** [2] and **Marc Aerts** [2]

[1] *Biostatistics Unit, South African Medical Research Council, Durban, South Africa*
[2] *I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium*
[3] *I-BioStat, KU Leuven, B-3000 Leuven, Belgium*

Outline

- Introduction
- Review of current approaches
- Proposed approach
- Application
- Discussion

Introduction

- Biomarkers for clinical events are particularly useful in the study of the HIV
- Two key biomarkers in HIV

  CD4 count (ARV initiation)

  Viral load (Treatment success)
- The time to reach a relevant CD4 count threshold in follow-up studies is used as a surrogate endpoint in studies examining HIV progression
- CD4 count subject to a high degree of fluctuation and measurement error
- A single CD4 count below a relevant threshold should be interpreted with caution
- Persistence criteria- two consecutive rule.
- Convention : First extract the time of the event, which is analysed in a second stage within the survival analysis framework.

Issues with the standard approach

- The standard approach assumes that the event times are observed without error
- Is not viable when the interval between visits is large
- Patients who enter the study with a CD4 count below the threshold are generally omitted - biases
- A method which takes into account the underlying marker trajectory, measurement error and left censoring is needed.

Model based approaches

- Multistate models
- Inverse prediction

  Extract the “true” patient specific marker trajectory

$$y_i = \beta_{oi} + \beta_{1i} t + \varepsilon_i$$

$$T_i = \frac{k - \beta_{oi}}{\beta_{1i}}$$

Issues:

- Cannot accommodate complex functions of time
- In the classical framework the properties of T_i are difficult to compute and simulation may be necessary

Proposed approach

- Stage 1 : A mixed model is fitted to the longitudinal measurements, resulting in patient-specific predicted values which are a function of the fixed-effects and empirical Bayes estimates.

- Stage 2: The probability of experiencing two consecutive measurements less than a relevant threshold k at each time point is computed. Using these estimates, the time to obtain two consecutive low CD4 counts is computed.

Methodology

Letting Y_{ij} denote the CD4 count observed on individual i at time point j, where $j = 1$ corresponds to an occasion at which one starts considering the individual as possibly seroconverting, the time to event T_i can be expressed as

$$T_i = \min\{j \geq 2 : Y_{ij-1} \leq k, Y_{ij} \leq k\}. \qquad (1)$$

It follows that the expected time for individual i to attain two consecutive CD4 counts less than the threshold k can be expressed as follows:

$$\begin{aligned} E(T_i) &= t_{i2} P(Y_{i1} \leq k, Y_{i2} \leq k) + t_{i3} P(Y_{i1} > k, Y_{i2} \leq k, Y_{i3} \leq k) \\ &\quad + t_{i4} \left\{ \begin{array}{l} P(Y_{i1} > k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k) \\ + P(Y_{i1} \leq k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k) \end{array} \right\} \\ &\quad + \cdots \\ &= \sum_{j=2}^{\infty} t_{ij} S_{ij}, \qquad (2) \end{aligned}$$

where S_{ij} denotes the probability of individual i experiencing the event, or 'stopping', at t_{ij}.

Methodology

We specify a linear mixed model which satisfies

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \qquad (3)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D),$$

$$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma_i),$$

where $\mathbf{b}_1, \ldots, \mathbf{b}_N, \boldsymbol{\varepsilon}_i, \ldots, \boldsymbol{\varepsilon}_N$ are independent. $\boldsymbol{\beta}$ and $\mathbf{b}_i$ represent the fixed and random effects, respectively. It follows that

$$\mathbf{Y}_i|\mathbf{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \Sigma_i).$$

Methodology

Assuming conditional independence in the linear mixed model such that $\Sigma_i = \sigma^2 I_{n_i}$, the joint probabilities which form S_{ij} reduce to the product of the individual probabilities. Hence,

$$S_{ij}|X_i, Z_i, \mathbf{b}_i, \boldsymbol{\beta} = C_{ij-3} P(Y_{ij-2} > k) P(Y_{ij-1} \leq k) P(Y_{ij} \leq k)$$
$$= C_{ij-3}\left[1 - \widetilde{\Phi}_{ij-2}(k)\right]\left[\widetilde{\Phi}_{ij-1}(k)\right]\left[\widetilde{\Phi}_{ij}(k)\right],$$

where C_{ij-3} denotes the 'continuation probability' at time t_{ij-3} and $\widetilde{\Phi}_{ij}(k)$ is a cumulative normal distribution with mean $\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$ and variance σ^2.

Computational efficiency

- It follows that $\widetilde{\Phi}_{ij}(k)$ can be expressed as a simple function of the standard univariate normal distribution:

$$\widetilde{\Phi}_{ij}(k) = \Phi\left(\frac{k - \mathbf{x}'_{ij}\boldsymbol{\beta} - \mathbf{z}'_{ij}\mathbf{b}_i}{\sigma}\right)$$

- Recursive relationship of continuation probabilities

$$C_{ij} = C_{j-2}[1 - \Phi_{ij-1}(k)][\Phi_{ij}(k)] + C_{j-1}[1 - \Phi_{ij}(k)].$$

Estimation and Inference

We propose a conditional version of the non-parametric case bootstrap to compute 95% confidence intervals for $\hat{T}_i$ as follows:

- Step 1. Individual i is removed from the full dataset resulting in N - 1 cases
- Step 2. Sample N - 1 subjects with replacement from the dataset in Step 1
- Step 3. Append the data of individual i to the bootstrap sample
- Step 4. Compute $\hat{T}_i$

This process is repeated 1000 times.

Application

- The Sinikithemba cohort comprises 336 HIV-1 subtype C *chronically infected* adults enrolled in the McCord Hospital (Durban, South Africa) between August 2003 and 2008
- CD4 count and viral load were measured every 3 and 6 months, respectively, from enrollment.
- Guidelines implemented during the study period, patients were recommended to start ARV treatment upon reaching a CD4 count less than 200 cells/mm^3 or WHO stage 3 or 4 symptoms.
- The median CD4 count at enrolment was 357 (IQR: 259-509) cells/mm^3 and the mean viral load was 4.7 log copies/ml.

Application

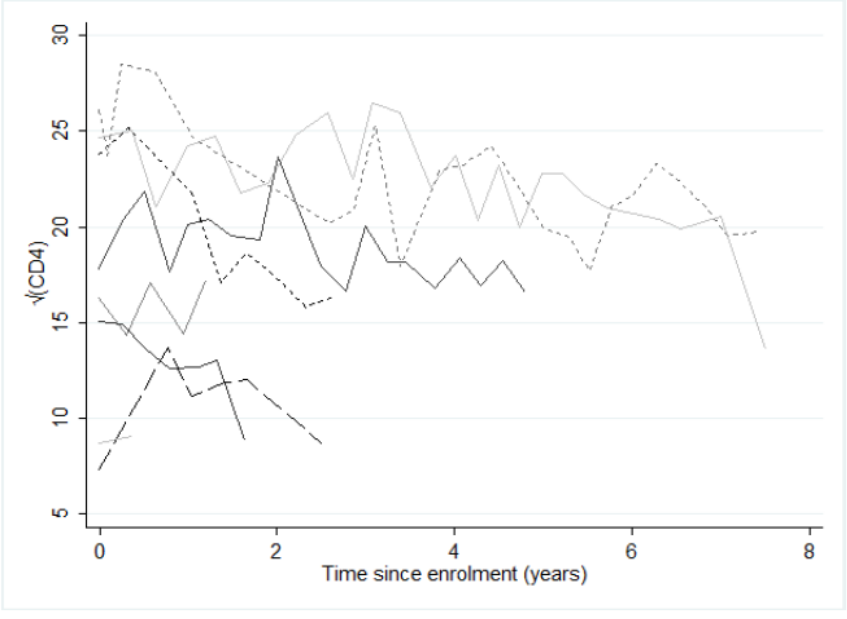


Figure 1: *Longitudinal CD4 count measurements for 8 subjects on the square root scale*

Stage 1 : Linear mixed model

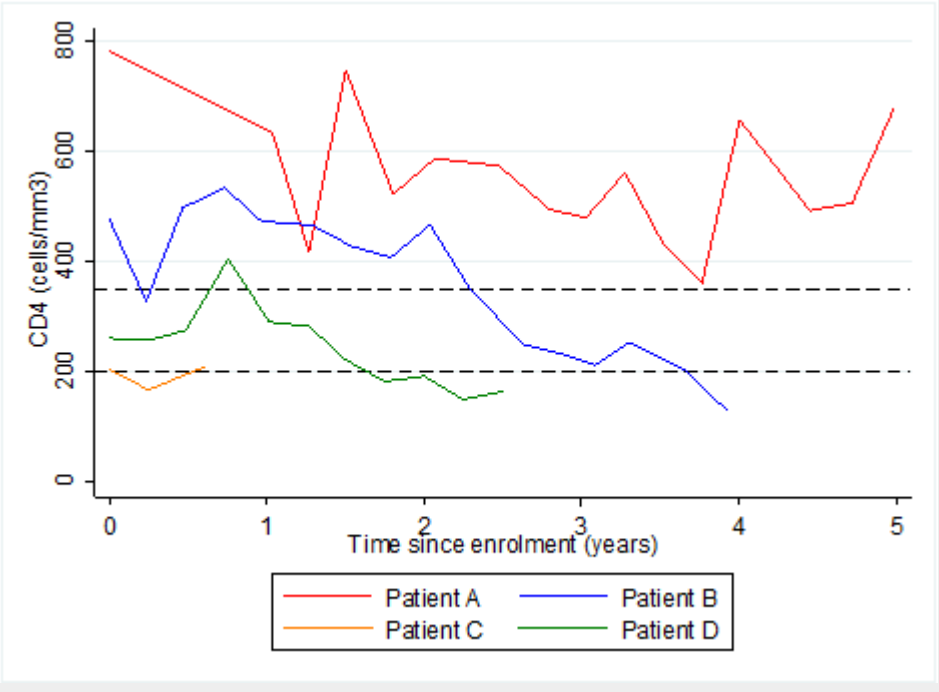
Table 1: *HIV cohort Data. Parameter estimates (standard errors) for the fitted models on each timescale*

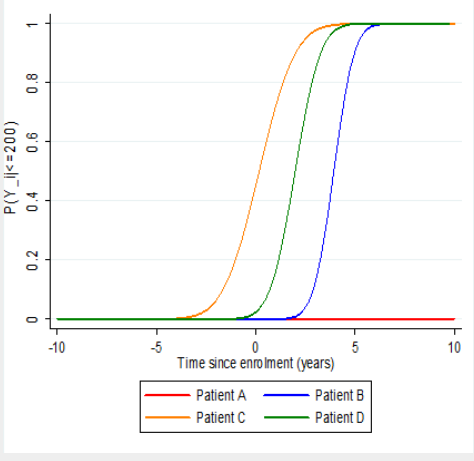
Effect	Parameter	Timescale enrolment origin	Timescale Calendar origin
Fixed effects estimates (s.e.)			
Intercept	$\beta_{0,L}$	21.2405 (0.471)	22.0000 (0.551)
	$\beta_{0,M}$	19.4469 (0.419)	20.6554 (0.498)
	$\beta_{0,H}$	16.2821 (0.406)	17.5021 (0.491)
Time	$\beta_{1,L}$	-0.5744 (0.121)	-0.5658 (0.117)
	$\beta_{1,M}$	-1.0160 (0.114)	-0.9454 (0.110)
	$\beta_{1,H}$	-1.3839 (0.140)	-1.1066 (0.133)
Covariance parameter estimates (s.e.)			
$var\ (b_{0i})$	d_{11}	19.5555 (1.608)	25.5456(2.272)
$cov\ (b_{0i}, b_{1i})$	d_{12}	-0.4944 (0.382)	-2.1611 (0.470)
$var\ (b_{1i})$	d_{22}	0.9941 (0.142)	0.9438 (0.130)
Measurement error	σ^2	3.1923 (0.081)	3.2135 (0.081)
Fit statistics			
AIC		17185.3	17225.9
BIC		17200.5	17241.1
-2 REML log-likelihood		17177.3	17217.9

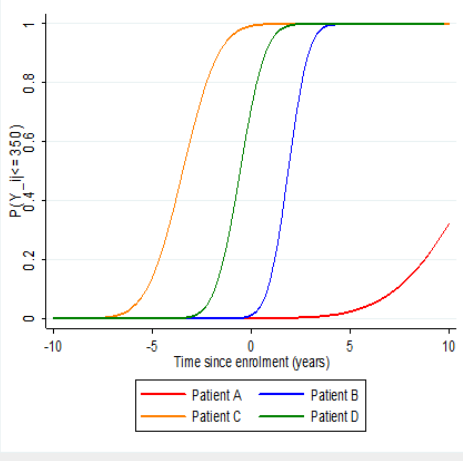
Assumptions

- We allow a 10-year window relative to enrolment where we consider an individual as having the potential to have experienced the threshold.
- The rationale for this decision is based on the estimated time from seroconversion to death in ART naive patients which was reported to be approximately 10 years in Sub-Saharan Africa.
- The discrete times which fall outside of the observation period were created in accordance with the study design of three monthly visits.
- The series was truncated at the visit at which $\hat{Y}_{ij}$ dropped to zero. Similarly, time t_{i1} was defined as the minimum time at which $\hat{Y}_{ij}$ < 1500 cells/mm^3,

Stage 2: Predicted probabilities



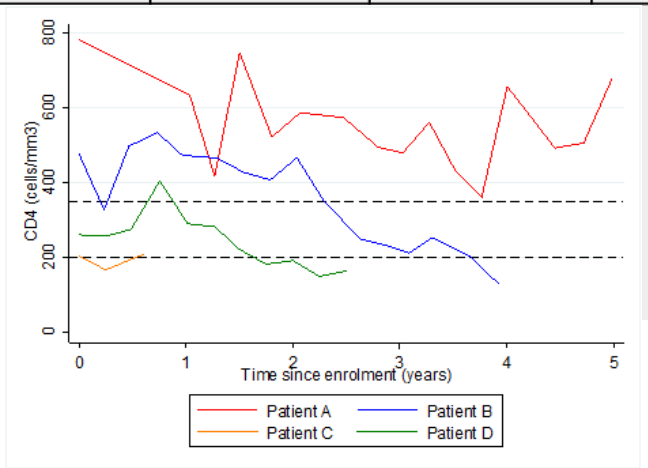




Stage 2: Estimated time to threshold

Table 2: *Estimated time to threshold for patients A, B, C, and D*

			≤ 200 cells/mm^3		≤ 350 cells/mm^3	
Patient	VL	Baseline CD4	$\widehat{T}_i$	95% CI	$\widehat{T}_i$	95% CI
A	Low	783	2.92 x 10−5	(8.88 x 10^{-6} , 8.24 x 10^{-5})	3.1552	(2.4946, 3.7362)
B	Low	478	4.2858	(4.2343, 4.3843)	2.3046	(2.2887, 2.3169)
C	High	204	0.3758	(0.0267, 0.5319)	-3.2608	(-5.0051, -2.2874)
D	High	261	2.3335	(2.3005, 2.3763)	-0.2043	(-0.4039, -0.0642)



Key findings

- 30 individuals had a zero probability of obtaining a CD4 count<200 throughout the period considered – Long term non-progressors?

Excluding the individuals who were long term non-progressors, the percentiles of the estimated times were computed.

- 15% of these patients had already attained two consecutive CD4 counts less than 200 more than six months prior to first presentation at the clinic.

- 35% of patients had already attained two consecutive CD4 counts less than 350 cells/mm^3 more than two years prior to enrollment.

Sensitivity analysis

Scenario 1. A period of 10 years prior to and post enrolment was considered, and visits outside the observed period occurred at regular three monthly intervals.

Scenario 2. A period of 5 years prior to and post enrolment was considered. Visits outside the observed period occurred at regular three monthly intervals.

Scenario 3. A period of 10 years prior to and post enrolment was considered and 10% of visits outside the observation period occurred one month later than expected.

Scenario 4. A period of 10 years prior to and post enrolment was considered and 25% of visits outside the observation period occurred one month later than expected.

Scenario 5. A period of 10 years prior to and post enrolment was considered and 10% of visits outside the observation period were missed.

Scenario 6. A period of 10 years prior to and post enrolment was considered and 20% of visits outside the observation period were missed.

Sensitivity analysis : Results

Table 3: *Estimated time to two consecutive measurements less than 350 cells/mm3 under various scenarios*

Patient	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
A	3.1552	0.0050	3.0734	3.1271	3.0219	2.5941
B	2.3046	2.3046	2.2926	2.2926	2.2926	2.2926
C	-3.2608	-3.2056	-3.2056	-3.2073	-3.2119	-2.9572
D	-0.2043	-0.2043	-0.2030	-0.2097	-0.1432	0.0208

- Exercise caution when interpreting estimated times to threshold in patients with very slow decline
- Methodology appears robust for the “general” patient

Other areas of application

- Diabetes
- Prostate cancer
- Abnormal aortic aneurysms

Conclusions and further work

- Methodology proposed is flexible and computationally efficient
- Additional sensitivity analysis is required

  Drop-out (MNAR?)
- Extension to accommodate correlated residuals
- Different stopping rules

References

WHO (2015). Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. *World Health Organization*

Sweeting, M. and Thompson, S. (2012). Making predictions from complex longitudinal data, with application to planning monitoring intervals in a national screening programme. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* **11,** 569–586.

Mandel, M. (2010). Estimating disease progression using panel data. *Biostatistics* **175,** 304–316.

Verbeke, G. and Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data*. New York: Springer.