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

PSI Journal Club – 22 March 2017

Tarylee Reddy¹,², Geert Molenberghs²,³, Edmund Njeru Njagi² and Marc Aerts²

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

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

$$E(T_i) = t_{i2}P(Y_{i1} \leq k, Y_{i2} \leq k) + t_{i3}P(Y_{i1} > k, Y_{i2} \leq k, Y_{i3} \leq k)$$

$$+ t_{i4}\left\{P(Y_{i1} > k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k)\right\}$$

$$+ \cdots$$

$$= \sum_{j=2}^{\infty} t_{ij}S_{ij},$$  \hfill (2)

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

\[ Y_i = X_i \beta + Z_i b_i + \varepsilon_i \quad (3) \]

\[ b_i \sim N(0, D), \]

\[ \varepsilon_i \sim N(0, \Sigma_i), \]

where \( b_1, \ldots, b_N, \varepsilon_i, \ldots, \varepsilon_N \) are independent. \( \beta \) and \( b_i \) represent the fixed and random effects, respectively. It follows that

\[ Y_i | b_i \sim N(X_i \beta + Z_i b_i, \Sigma_i). \]
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,

$$S_{ij}|X_i, Z_i, b_i, \beta = C_{ij-3} P(Y_{ij-2} > k)P(Y_{ij-1} \leq k)P(Y_{ij} \leq k)$$

$$= C_{ij-3} [1 - \Phi_{ij-2}(k)][\Phi_{ij-1}(k)][\Phi_{ij}(k)],$$

where $C_{ij-3}$ denotes the ‘continuation probability’ at time $t_{ij-3}$ and $\Phi_{ij}(k)$ is a cumulative normal distribution with mean $x_{ij}^T \beta + z_{ij}^T b_i$ and variance $\sigma^2$. 
Computational efficiency

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

$$
\tilde{\Phi}_{ij}(k) = \Phi \left( \frac{k - x_{ij}'\beta - z_{ij}'b_i}{\sigma} \right)
$$

• Recursive relationship of continuation probabilities

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

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

<table>
<thead>
<tr>
<th>Patient</th>
<th>VL</th>
<th>Baseline CD4</th>
<th>$\hat{T}_i$</th>
<th>95% CI</th>
<th>$\hat{T}_i$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Low</td>
<td>783</td>
<td>2.92 x 10⁻⁵</td>
<td>(8.88 x 10⁻⁶, 8.24 x 10⁻⁵)</td>
<td>3.1552</td>
<td>(2.4946, 3.7362)</td>
</tr>
<tr>
<td>B</td>
<td>Low</td>
<td>478</td>
<td>4.2858</td>
<td>(4.2343, 4.3843)</td>
<td>2.3046</td>
<td>(2.2887, 2.3169)</td>
</tr>
<tr>
<td>C</td>
<td>High</td>
<td>204</td>
<td>0.3758</td>
<td>(0.0267, 0.5319)</td>
<td>-3.2608</td>
<td>(-5.0051, -2.2874)</td>
</tr>
<tr>
<td>D</td>
<td>High</td>
<td>261</td>
<td>2.3335</td>
<td>(2.3005, 2.3763)</td>
<td>-0.2043</td>
<td>(-0.4039, -0.0642)</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Patient</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
<th>Scenario 5</th>
<th>Scenario 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3.1552</td>
<td>0.0050</td>
<td>3.0734</td>
<td>3.1271</td>
<td>3.0219</td>
<td>2.5941</td>
</tr>
<tr>
<td>B</td>
<td>2.3046</td>
<td>2.3046</td>
<td>2.2926</td>
<td>2.2926</td>
<td>2.2926</td>
<td>2.2926</td>
</tr>
<tr>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3.2608</td>
<td>3.2056</td>
<td>3.2056</td>
<td>3.2073</td>
<td>3.2119</td>
<td>2.9572</td>
</tr>
<tr>
<td>D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0.2043</td>
<td>0.2043</td>
<td>0.2030</td>
<td>0.2097</td>
<td>0.1432</td>
<td>0.0208</td>
</tr>
</tbody>
</table>

- 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*

