Joint Analysis of Longitudinal and Survival AIDS Data with a Spatial Fraction of Long-term Survivors: A Bayesian Approach

Rui Martins¹, Giovanni Silva²,⁴ & Valeska Andreozzi³,⁴

¹Universidade de Lisboa; ²Egas Moniz - Cooperativa de Ensino Superior, CRL.
³Centro de Investigação Interdisciplinar Egas Moniz (CIME), Escola Superior de Saúde Egas Moniz (SSEEM); ⁴Instituto Superior Técnico, Universidade de Lisboa; (¹Faculdade de Ciências da Universidade Nova de Lisboa; (2)Centro de Estatística e Aplicação de Universidade de Lisboa (CEAUL).

Overview

The objective is to model longitudinal and survival data jointly taking into account the dependence between the two responses in a real HIV/AIDS dataset. We employ here a Bayesian hierarchical approach to jointly model spatially-clustered survival data with a fraction of long-term survivors along with longitudinal measurements of CD4⁺ T lymphocyte counts for a random sample of 500 HIV/AIDS individuals collected in all the 27 states of Brazil during the period 2002–2006.

In order to accommodate a more flexible choice for the longitudinal model we propose to use a specification via penalized B-Splines [2]. To deal with the possible fact that only a subpopulation undergoes the event we propose a spatial cure model, which helps in explaining the behavior of chronic or potentially terminal diseases in different regions. Classical survival models (e.g. Cox model) are not well suited to take into account long-term survivors.

Results show that using a cure fraction allows us to improve the results comparatively to a survival analysis without it. The inclusion of spatial frailties eases in mapping the heterogeneity in the risk among the Brazilian states and helps in the explanation of the hazard.

Dataset

- Data origin: Brazilian database on HIV;
- Period (years): 2002–2006;
- Sample size: n = 500 individuals
- Response variables: y = CD4⁺ T lymphocyte counts and survival time (years since the patient’s entry in the study until death)
- Explanatory variables: age (<50=0, ≥50=1); gender (Female=0, Male=1); PrevOI (previous opportunistic infection at study entry=1, no previous infection=0); region of residence (one of the 27 Brazilian States);
- 34 deaths. 88% of the patients were between 15 and 49 years old; 60% were males.

The CD4 counts initial median was 245 cells/mm³ (men - 226 cells/mm³; women - 263 cells/mm³).

References


Model

Consider the CD4 repeated measurements, y_{ik} = (y_{ik1}, ..., y_{ikq}), and the observed (possibly right censored) time-to-death, T_{ik}, for the i-th individual living in the k-th region, k = 1, ..., K, i = 1, ..., n_k. Longitudinal data is described by a nonlinear mixed effects model,

\begin{equation}
\begin{align*}
\text{y}_{ikj} | \text{b}_{ik}, \sigma_j^2 &\sim \mathcal{N}(\text{y}_{ikj}(\text{t}_{ikj}), \sigma_j^2), \quad j = 1, ..., p_{ik}, \\
\text{y}_{ikj}(\text{t}_{ik}) &\sim (\beta_{i0} + \text{b}_{ijk} + m_{ip_{ik}j}\text{t}_{ikj}; \beta_2 + m_{ip_{ik}2}\text{t}_{ikj}; \text{b}_{ik}),
\end{align*}
\end{equation}

where \(m_{ip_{ik}j}\) and \(m_{ip_{ik}2}\) are the overall and individual nonlinear effects, respectively, considered here as Penalized cubic B-Splines. \(\beta_{i0}\) and \(\beta_{i1}\) are the common and subject-specific intercept, respectively, \(\beta_k = (\beta_{k1}, ..., \beta_{kp_{ik}})^T\), whose elements are assumed to be mutually independent, represents the fixed effects of the subject-specific vector of covariates, \(\text{z}_{k6k} = (\text{z}_{k61k}, ..., \text{z}_{k6p_{ik}})^T\) and \(\text{b}_{ik} = (\text{b}_{ik1}, ..., \text{b}_{ikp_{ik}})^T\) are, respectively, population and individual-specific regression parameters for each basis function, penalized through a second order random-walk. Assuming the time-to-death of the non-cured group is Weibull distributed, \(W(\rho, \nu(0))\), the spatial cure model [1] is described as,

\begin{equation}
S_p(t) = (1 - \theta_k) + \theta_k S_{ik}(t) = (1 - \theta_k) + \theta_k \exp\left\{-\rho \nu(0)\right\},
\end{equation}

being \(S_p\) and \(S_{ik}\) the survival functions for the entire population and for the non-cured group, respectively. \((1 - \theta_k)\) is a region-specific cure fraction. The baseline covariates and the longitudinal information will be introduced through the scale parameter, \(\nu(0)\), allowing it to vary across individuals and regions:

\begin{equation}
\nu_q(t) = \text{x}_{ik}^T\beta_q + \gamma_{ik}^T\theta_{ik}(t) + W_k,
\end{equation}

where \(\gamma\) is a parameter quantifying the effect of the CD4 values to the survival; \(\text{a}_{ik}\) is a vector of baseline covariates (can coincide with \(\text{z}_{ik}\)), \(\beta_q\) is the respective vector of coefficients and \(W_k\) is a region-specific frailty, \(W_k/\nu_q^{-1} \sim I^\nu Q(CV)\).

Application

Several scenarios for the Penalized splines were tried. Namely the internal knots were fixed every: 1 month, 2 months, 3 months, 6 months and 12 months providing, respectively, 59, 29, 11, 9, 4 and internal knots. The best fit was always achieved with 19 internal knots during the 5 years, resulting in \(Q = 23\) cubic basic functions.

\begin{equation}
\begin{align*}
y_{ik} &= (\beta_{i0} + \text{b}_{ijk}) + \beta_{i1}\text{sex}_{ik} + \beta_{i2}\text{age}_{ik} + \beta_{i3}\text{PrevOL}_{ik} + \sum_{q=1}^{Q-3} \beta_{iq}\text{B}_{iq}(t_{ik}) + \sum_{q=1}^{Q-3} \beta_{iq2}\text{B}_{iq2}(t_{ik}) \quad (5) \\
\nu_{ik}(t) &= \text{b}_{i11} + \beta_{i22}\text{sex}_{ik} + \beta_{i33}\text{age}_{ik} + \beta_{i34}\text{PrevOL}_{ik} + \gamma \nu(t) + W_k
\end{align*}
\end{equation}

Results

<table>
<thead>
<tr>
<th>((\beta_{i0}))</th>
<th>((\beta_{i1})) sex</th>
<th>((\beta_{i2})) age</th>
<th>((\beta_{i3})) PrevOL</th>
<th>((\sigma_q^2))</th>
<th>((\gamma))</th>
<th>((\nu_q(0)))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
</tr>
<tr>
<td>14.60</td>
<td>(8.37, 18.78)</td>
<td>-0.15</td>
<td>(-2.66, 2.08)</td>
<td>-1.30</td>
<td>(-4.25, 1.30)</td>
<td>-1.84</td>
</tr>
</tbody>
</table>
\(\nu_q(0)\) | 0.015 | (0.0002, 0.13) | 2.34 | (3.67, 3.11) | 0.17 | (0.01, 0.47) |

Model choice: Watanabe-Akaike Information Criterion (WAIC), Logarithm of the Pseudo-Marginal Likelihood (LPML).

Spatial frailties: Brazil’s map presenting the posterior median of the region-specific relative risks, \(\exp(\nu_{ik}(0))\).

Acknowledgments

The authors wish to thank Maria Gorieta Fonseca and Claudia Coeli by the database. This work was partly supported by Fundação para a Ciência e a Tecnologia (FCT) projects PTDC/MAT/118355/2010 and UID/MAT/00006/2013 and Egas Moniz - Cooperativa de Ensino Superior, CRL.

The model choice measures, WAIC and LPML, agree in the the selected model (SM) ⇒ Considering a common cure fraction for all patients and a spatial frailty outperforms the traditional joint model approach for the HIV/AIDS data, which does not consider a fraction of long-term survivors. The estimated cure fraction, \((1 - \theta_k)\), indicates that nearly 17% of the HIV/AIDS patients in this study may be considered long-term survivors.