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

Rui Martins[1] , **Giovani Silva**[2,4] & **Valeska Andreozzi**[3,4]

(1)c.a. ruimartins@egasmoniz.edu.pt, Centro de Investigação Interdisciplinar Egas Moniz (CiiEM), Escola Superior de Saúde Egas Moniz (ESSEM); (2)Instituto Superior Técnico, Universidade de Lisboa; (3)Faculdade de Ciências da Universidade Nova de Lisboa; (4)Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL)

CENTRO DE INVESTIGAÇÃO INTERDISCIPLINAR EGAS MONIZ

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 spatial-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**: $y = \sqrt{\text{CD4}^+\text{T}}$ 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/mm^3 (men - 226 cells/mm^3; women - 263 cells/mm^3).

References

[1] Cooner, F., Banerjee, S., and McBean, A.M. (2006) Modelling geographically referenced survival data with a cure fraction. *Stat Meth Med Res*, **15**(4):307–324;

[2] Lang, S., Brezger, A. (2004) Bayesian p-splines. *J Comp Graph Stat*, **13**(1):183–212;

[3] Martins, R., Silva, G.L., and Andreozzi, V. (2016) Bayesian joint modeling of longitudinal and spatial survival AIDS data. *Stat Med*, doi: 10.1002/sim.6937.

[4] Rizopoulos, D., Ghosh, P. (2011) A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat Med*, **30**(12):1366–1380;

Acknowledgments

The authors wish to thank Maria Goretti Fonseca and Cláudia Coeli by the database. This work was partially funded by Fundação para a Ciência e a Tecnologia (FCT) projects PTDC/MAT/118335/2010 and UID/MAT/00006/2013 and Egas Moniz - Cooperativa de Ensino Superior, CRL.

Model

Consider the CD4 repeated measurements, $\mathbf{y}_{ik} = (y_{ik1}, \ldots, y_{in_{ik}})$, and the observed (possibly right censored) time-to-death, T_{ik}, for the ith individual living in the kth region, $k = 1, \ldots, K$, $i = 1, \ldots, n_k$. **Longitudinal data** is described by a **nonlinear mixed effects model**,

$$y_{ikj}|\boldsymbol{b}_{ik}, \sigma_e^2 \sim \mathcal{N}(y^*_{ik}(t_{ikj}), \sigma_e^2), \quad j = 1, \ldots, n_{ik}, \tag{1}$$

$$y^*_{ik}(t_{ikj}) = (\beta_{10} + b_{ik0}) + \mathbf{z}_{ik}^\top \boldsymbol{\beta}_1 + m_p(t_{ikj}; \boldsymbol{\beta_2}) + m_{ik}(t_{ikj}; \boldsymbol{b}_{ik}), \tag{2}$$

where $m_p(.)$ and $m_{ik}(.)$ are the overall and individual nonlinear effects, respectively, considered here as **Penalized cubic B-Splines**. β_{10} and b_{ik0} are the common and subject-specific intercept, respectively. $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1r})^\top$, whose elements are assumed to be mutually independent, represents the fixed effects of the subject-specific vector of covariates, $\mathbf{z}_{ik}$; $\boldsymbol{\beta}_2 = (\beta_{21}, \ldots, \beta_{2Q})^\top$ and $\boldsymbol{b}_{ik} = (b_{ik1}, \ldots, b_{ikQ})^\top$ are, respectively, population and individual-specific **regression parameters for each basis function, penalized through a second order random-walk**.
Assuming the time-to-death for the non-cured group is Weibull distributed, $\mathcal{W}(\rho, e^{\eta(t)})$, the **spatial cure model** [1] is described as,

$$S_p(t) = (1 - \theta_k) + \theta_k S(t) = (1 - \theta_k) + \theta_k \exp\left\{-t^\rho e^{\eta(t)}\right\}, \tag{3}$$

being S_p and S 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, $e^{\eta(t)}$, allowing it to **vary across individuals and regions**:

$$\eta_{ik}(t) = \mathbf{x}_{ik}^\top \boldsymbol{\beta}_3 + \gamma\, y^*_{ik}(t) + W_k, \tag{4}$$

where γ is a parameter quantifying the effect of the CD4 values to the survival; $\mathbf{x}_{ik}$ is a vector of baseline covariates (can coincide with $\mathbf{z}_{ik}$), $\boldsymbol{\beta}_3$ is the respective vector of coefficients and W_k is a region-specific frailty, $W_k|\sigma^2_W \sim ICAR(\sigma^2_W)$.

Application

Several scenarios for the Penalized splines were tried. Namely the internal knots were fixed every: 1 month, 2 months, 3 months, 4 months, 6 months and 12 months providing, respectively, 59, 29, 19, 14, 9 and 4 internal knots. The best fit was always achieved with 19 internal knots, corresponding roughly to a knot every 3 months during the 5 years, resulting in $Q = 23$ cubic basis functions.

$$y^*_{ikj} = (\beta_{10} + b_{ik0}) + \beta_{11}\text{sex}_{ik} + \beta_{12}\text{age}_{ik} + \beta_{13}\text{PrevOI}_{ik} + \sum_{q=1}^{Q=23} \beta_{2q} B_q(t_{ikj}) + \sum_{q=1}^{Q=23} b_{ikq} B_q(t_{ikj}) \tag{5}$$

$$\eta_{ik}(t) = \beta_{31} + \beta_{32}\text{sex}_{ik} + \beta_{33}\text{age}_{ik} + \beta_{34}\text{PrevOI}_{ik} + \gamma\, y^*_{ik}(t) + W_k \tag{6}$$

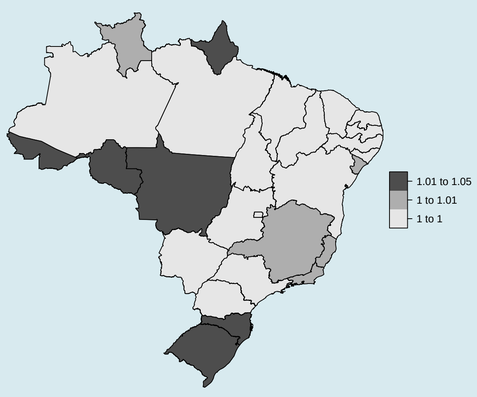
Results

$\pi(\theta)$	W_k	Measure
Common Cure	0	14867
		-7473
$\theta_k = \theta, \forall k$	W_k	**14540**(SM)
$\theta \sim \mathcal{U}(0,1)$		**-7453**(SM)
Regional Cure	0	14578
		-7482
$\theta_k \neq \theta_l, \forall k \neq l$	W_k	14596
$\theta_k \sim \mathcal{U}(0,1)$		-7487
	0	14575
No Cure		-7461
$1 - \theta_k = 0, \forall k$	W_k	14577
		-7496

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

	Mean	95% CI
(β_{10})	14.65	$(8.37, 18.78)$
(β_{11}) sex	-0.15	$(-2.66, 2.08)$
(β_{12}) age	-1.30	$(-4.25, 1.30)$
(β_{13}) PrevOI	-1.84	$(-3.67, -0.01)$
σ_e^2	4.99	$(4.54, 5.26)$
(β_{31})	-3.14	$(-4.42, -1.87)$
(β_{32}) sex	0.47	$(-0.35, 1.36)$
(β_{33}) age	0.58	$(-0.45, 1.61)$
(β_{34}) PrevOI	0.67	$(-0.15, 1.49)$
γ	-0.17	$(-0.25, -0.11)$
σ^2_W	0.015	$(0.0002, 0.13)$
ρ	2.34	$(1.67, 3.11)$
$1 - \theta$	0.17	$(0.01, 0.47)$

Parameters: posterior mean and its 95% Credibility Interval (CI).



Spatial frailties - Brazil's map presenting the posterior median of the region-specific relative risks, $\exp\{W_k\}$.

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 considers a fraction of long-term survivors. The estimated cure fraction, $(1 - \theta)$, indicates that nearly 17% of the HIV/AIDS patients in this study may be considered long-term survivors.