A linear mixed model for segmented regression with smooth transition

Julio M. Singer[1], Francisco M.M. Rocha[2], Antonio Carlos Pedroso-de-Lima[1], Giovani, L. Silva[3], Giuliana C. Coatti[4], e Mayana Zatz[4]

[1]Department of Statistics, University of Sao Paulo

[2]Paulista School of Politics, Economics and Business, Federal University of Sao Paulo

[3]Department of Mathematics, University of Lisbon

[4]Institute of Biosciences, University of Sao Paulo

Abstract

We consider random changepoint mixed segmented regression models to analyse data obtained from a study conducted to verify whether treament with stem cells may delay the onset of a symptom of amyotrophic lateral sclerosis in genetically modified mice. The proposed models capture the biological aspects of the data, accommodating a smooth transition between the periods with and without symptoms. An additional changepoint is considered to avoid negative predicted responses. Given the non-linear nature of the model, we adapt an algorithm proposed by Muggeo et al. (2014, Statistical Modelling) to estimate the fixed parameters and to predict the random effects by fitting linear mixed models at each step.

Key words

amyotrophic lateral sclerosis; fitting algorithm; mixed models; random effects.

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is one of the most common adult-onset motor neuron disease causing a progressive, rapid and irreversible degeneration of motor neurons in the cortex, brain stem and spinal cord. In the majority of cases ALS occurs sporadically; in about 10% of the cases it is caused by familial reasons. No effective treatment is available and cell therapy clinical trials are currently being tested in ALS affected patients. The SOD1 gene encodes an important antioxidant human enzyme and mutations in SOD1 represent one of the most frequent causes of ALS.

Among the different animal models for ALS, SOD1 mice are the most used in pre-clinical studies. After the initial tremor in the limbs they develop muscle weakness in early adulthood, become fully paralyzed and die. These mice over-express the human SOD1 gene bearing the G93A mutation, a point mutation found in familial ALS. Interestingly, in this animal model the disease progression is different between the genders. Males have a shorter lifespan and a clinical condition apparently more severe than females and differences

in electrophysiological parameters have also been reported. A comparable effect of gender is also observed in ALS patients.

Treatment of ALS with stem cells is a current research topic. Mesenchymal stromal cells (MSC), specially those derived from adipose tissues, and pericytes have been used in studies that focus on the reduction of the speed of the progression of symptoms of neuro-degenerative diseases. In this context we consider a study conducted in the Human Genome and Stem Cell Research Center, at the Biosciences Institute, University of S˜ao Paulo, Brazil with the objective of comparing MSC cells and pericytes injected in SOD1-G93A mice with respect to their effects on the evolution of some symptoms of ALS. Details may be obtained in Coatti et al. (2017).

Our objective here is to propose models for the statistical analysis of the data.

2. The study

A set of 34 female and 21 male 8 week old SOD1-G93A mice was divided into 3 groups Animals in the first group (12 females and 7 males) were submitted to weekly injections of MSC cells, those in second group (11 females and 8 males), to injection with pericytes while animals in the third group (11 females and 6 males) were submitted to the vehicle (*Hank's balanced salt solution* - HBSS). All animals were followed weekly up to their death for clinical analysis of the progression of the disease by means of four variables, the analysis of one of them, *rotarod* is considered in this study. The *rotarod* test was used to evaluate motor coordination and fatigue resistance. For that purpose, the length of time each animal could remain on the rotating cylinder (3.5 cm) of a *rotarod* apparatus (IITC Life Science model 755) was recorded. The initial speed was 1 rpm and it was increased constantly until a final speed of 30 rpm, after 180 s. Each animal was given three tries and the longest latency to fall was recorded. The specific objectives of the analysis are:

i) Identification of the moment when animals become symptomatic (symptoms onset) forthe six groups defined by the combination of treatment (HBSS, MSCs, pericytes) and sex (male, female).
ii) Estimation of the expected rate of variation in response after symptom onset for eachgroup.
iii) Evaluation of the effects of treatment, sex and their interaction on the expected momentof symptom onset and post onset rate of variation in the expected response.

3. Statistical analysis

Profile plots for the response along with LOESS curves are displayed in Figure 1.

A longitudinal analysis of the behaviour of the response variable corroborates its expected stable level before the onset of the symptom (a decrease in the length of time during which the animal remains in the rotating cylinder). Furthermore, individual differences in the moment where this occurs as well as differences among the accelerations with which the intensity of the symptom progresses are also visible. It also seems reasonable to expect a change in the acceleration with which the intensity of the symptom progresses after the disease onset.

Given that such conclusions are in line with the expected biological behaviour, a random changepoint mixed polynomial segmented regression model may be considered for the analysis.

Such models have an attractive practical appeal in many fields and have been the object of statistical research for a long time as detailed in Muggeo et al. (2014). These authors consider a frequentist approach as opposed to the commonly Bayesian perspective usually employed in the statistical literature.

Keeping in mind the necessarily non-negative nature of the response, we adopt a similar approach and consider an analysis of the ALS data based on the model

$$y_{ijk} = \alpha_{ij} I(t_k < \psi_{2ij}) + \gamma_{ij}[t_k - \psi_{1ij}(\lambda_{ij})]^2 I(\psi_{1ij} \leq t_k < \psi_{2ij}) + e_{ijk} \qquad (1)$$

($i = 1,...,6$, $j = 1,...,n_i$ and $k = 1,...,n_{ij}$) where y_{ijk} denotes the response for the j-th animal observed in the i-th group (defined by the combination of the levels of treatment and sex) at the k-th evaluation instant, α_{ij} is the corresponding stable level of the symptom prior to the first changepoint, γ_{ij} is the coefficient of the quadratic term for the curve that governs the response behaviour post changepoint ψ_{1ij}, with

$$\psi_{1ij}(\lambda_{ij}) = [L_1 + L_2 \exp(\lambda_{ij})]/[1 + \exp(\lambda_{ij})]$$

to restrict the value of ψ_{1ij} to the interval (L_1, L_2) in which the observations are obtained and ψ_{2ij} denotes the instant where the response is null. We assume that $\alpha_{ij} = \alpha_i + a_{ij}$, $\gamma_{ij} = \gamma_i + c_{ij}$, $\lambda_{ij} = \lambda_i + \ell_{ij}$ with $\mathbf{b}_{ij} = (a_{ij}, c_{ij}, \ell_{ij})^\top \sim N(\mathbf{0}, \mathbf{G}_i)$ and $e_{ijk} \sim N(0, \sigma_i^2)$ independent of $\mathbf{b}_{ij}$.

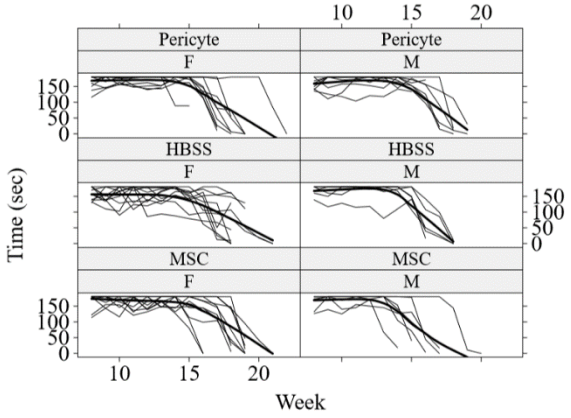


Figure 1: Profile plots for the response along with LOESS curves

This is an extension of the models proposed by Muggeo et al. (2014) where a smooth transition and a second changepoint are incorporated. For the sake of notational simplicity and without loss of generality, we drop the subscript i to specify the the fitting algorithm.

Given that ψ_{2j} corresponds to the instant t_k where $y_{jk} = 0$, we have $I(t_k < \psi_{2j}) = 1$ and $I(\psi_{1j} \leq t_k < \psi_{2j}) = 1$ and consequently, that $\alpha_j + \{\gamma_j[\psi_{2j} - \psi_{1j}(\lambda_j)]^2\} = 0$, implying that

$$\psi_{2j} = \psi_{2j}(\alpha_j, \gamma_j, \psi_{1j}) = \sqrt{-\alpha_j/\gamma_j} + \psi_{1j}(\lambda_j)$$

Following Muggeo et al. (2014) and Fasola et al. (2018), the model, which is non-linear, may be approximated by a first order Taylor expansion of $f[t_k, \gamma_j, \psi_{1j}(\lambda_j)] = \gamma_j[t_k - \psi_{1j}(\lambda_j)]_2 I(\psi_{1j} \leq t_k < \psi_{2j})$.
Explicitly,

$$f[t_k, \gamma_j, \psi_{1j}(\lambda_j)] \approx f[t_k, \gamma_j, \psi_{1j}(\widehat{\lambda}_j)] + (\lambda_j - \widehat{\lambda}_j)\frac{\partial f[t_k, \gamma_j, \psi_{1j}]}{\partial \psi_{1j}}\frac{\partial \psi_{1j}(\lambda_j)}{\lambda_j}\Big|_{\lambda_j = \widehat{\lambda}_j}$$

with

$$\frac{\partial f[t_k, \gamma_j, \psi_{1j}]}{\partial \psi_{1j}} = h_j(\lambda_j) = 2\gamma_j[t_k - \psi_{1j}(\lambda_j)]I[\psi_{1j}(\lambda_j) \leq t_k < \psi_{2j}]$$

and

$$\frac{\partial \psi_{1j}(\lambda_j)}{\partial \lambda_j} = g_j(\lambda_j) = \frac{(L_2 - L_1)\exp(\lambda_j)}{[1 + \exp(\lambda_j)]^2}.$$

Consequently we may approximate model (1) by

$$y_{jk} \approx \alpha_j I(t_k < \psi_{2j}) + f[t_k, \gamma_j, \psi_{1j}(\widehat{\lambda}_j)] - \widehat{\lambda}_j h_j(\widehat{\lambda}_j) g_j(\widehat{\lambda}_j) + \lambda_j h_j(\widehat{\lambda}_j) g_j(\widehat{\lambda}_j) + e_{jk}. \quad (2)$$

Considering the pseudo observations defined by $y^*_{jk} = y_{jk} + \widehat{\lambda}_j h_j(\widehat{\lambda}_j) g_j(\widehat{\lambda}_j)$, the model

$$y^*_{jk} = \alpha_j I(t_k < \psi_{2j}) + f[t_k, \gamma_j, \psi_{1j}(\widehat{\lambda}_j)] + \lambda_j h_j(\widehat{\lambda}_j) g_j(\widehat{\lambda}_j) + e_{jk}$$

suggests the following algorithm to fit (1)

1) Let $\psi_{1j}^{(0)} = \psi_1^{(0)}$ and $\psi_{2j}^{(0)} = \psi_2^{(0)}$.

2) Fit model $y_{jk} = \alpha_j I(t_k < \psi_{2j}^0) + \gamma_j (t_k - \psi_{2j}^{(0)})^2 I(\psi_{1j}^{(0)} \leq t_k < \psi_{2j}^{(0)}) + e_{jk}$ to obtain $\alpha^{(0)}$, $a_j^{(0)}$, $\gamma^{(0)}$, $c_j^{(0)}$, $\lambda_j^{(0)} = \log[(\psi_{1j}^{(0)} - L_1)/(L_2 - \psi_{1j}^{(0)})]$ and $\psi_{2j}^{(1)} = \sqrt{-\alpha_j^{(0)}/\gamma_j^{(0)}} + \psi_{1j}^{(0)}$.

3) Let $r = 1$.

4) Compute $y_{jk}^{(r)} = y_{jk} + \lambda_j^{(r-1)} h_j(\lambda_j^{(r-1)}) g_j(\lambda_j^{(r-1)})$.

5) Fit model

$$y_{jk}^{(r)} = \alpha_j I(t_k < \psi_{2j}^{(r)}) + \gamma_j [t_k - \psi_{1j}^{(r)}]^2 I(\psi_{1j}^{(r)} \leq t_k < \psi_{2j}^{(r)}) + \lambda_j h_j(\lambda_j^{(r-1)}) g_j(\lambda_j^{(r-1)}) + e_{jk}^{(r-1)}$$

to obtain $\alpha^{(r)}$, $a_j^{(r)}$, $\gamma^{(r)}$, $c_j^{(r)}$, $\lambda^{(r)}$, $\ell_j^{(r)}$, $\psi_{1j}^{(r)} = [L_1 + L_2 \exp(\lambda_j^{(r)})]/[1 + \exp(\lambda_j^{(r)})]$ and $\psi_{2j}^{(r+1)} = \sqrt{-\alpha_j^{(r)}/\gamma_j^{(r)}} + \psi_{1j}^{(r)}$.

6) Stop if some convergence criterion is satisfied, otherwise, let $r = r + 1$ and repeat steps 4-6.

This algorithm, adapted from Muggeo et al. (2014), essentially considers iterative fitting of standard linear mixed models by restricted maximum likelihood. At convergence, we expect a neglible difference between the third and fourth terms in the right hand side of (2) and as a consequence, that the pseudo observations should well approximate the original ones. Given the linear mixed model nature of the proposed fitting algorithm, we may employ the diagnostic procedures outlined in Singer et al. (2017) to check whether the adopted assumptions for the distribution of the random effects or of the random error are reasonable.

Estimates of the parameters of model (1) obtained via fitting the approximation (2) along with the corresponding standard errors are summarized in Table (1).

The results of a Wald test for the homogeneity of the six changepoints ψ_1 ($\chi^2 = 58.30, df = 5, p < 0.001$) suggests further analyses to identify the possible effects of treatment, sex and their interaction. A significant interaction between treatment and sex with respect to the ψ_1 changepoints ($\chi^2 = 13.65, df = 2, p = 0.001$) may be analysed via the multiple comparisons summarized in Table 2 and suggest that the onset of symptoms for the control group (HBSS) males is delayed by 1.7 [CI(95%) = 1.0, 2.4] weeks with respect to the control group females and that treatment with Pericytes (both sexes) or MSC (females) delay the onset of symptoms by 1.3 [CI(95%) = 0.5, 2.2] weeks with respect to HBSS treated males. The changepoint for MSC treated males lies between those for HBSS treated males and females but the small sample size does not lead to a significant difference in either case.

The results for a similar analysis of the acceleration with which the symptom progresses suggest no difference between sexes and an increase in the acceleration of 18.6 [CI(95%) = 17.8, 19.6] sec/week^2 for the experimental

treatments (MSC and Pericytes) relatively to that of the control treatment (HBSS).

An example of predicted subject specific response curves is presented in Figure 2.

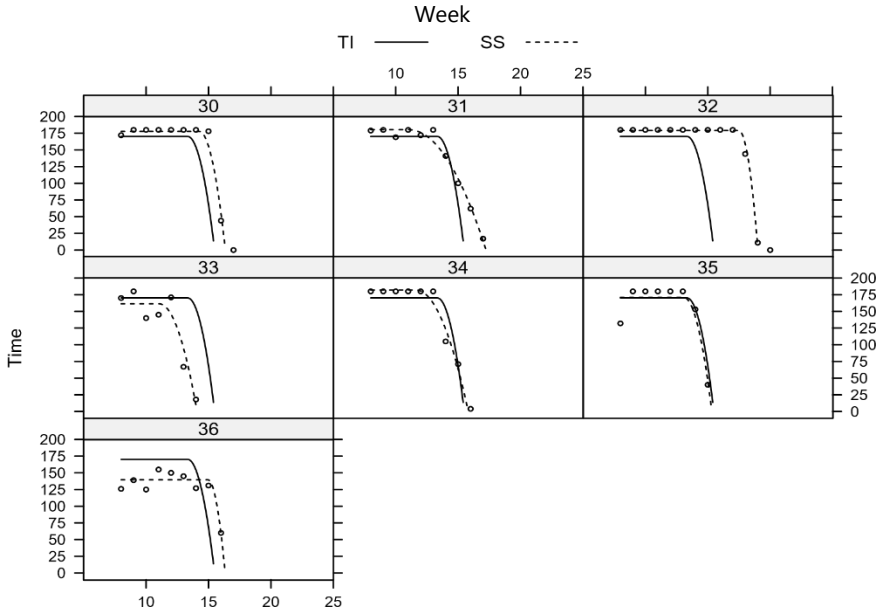
Parameter	Sex	Treatment	Estimate	Std error	
				Model	Robust
Intercept (α)	M	HBSS	166.5	10.0	9.1
	M	MSC	170.2	6.3	5.8
	M	Pericytes	164.5	6.5	6.1
Intercept (α)	F	HBSS	156.6	6.5	6.2
	F	MSC	167.4	3.8	3.6
	F	Pericytes	170.2	3.6	3.4
2nd degree coefficient (γ)	M	HBSS	-18.2	4.7	4.2
	M	MSC	-36.7	11.4	10.4
	M	Pericytes	-23.1	12.5	11.7
2nd degree coefficient (γ)	F	HBSS	-2.8	1.0	0.9
	F	MSC	-26.6	6.3	6.0
	F	Pericytes	-30.0	7.8	7.3
Changepoint 1 (ψ_1)	M	HBSS	13.9	0.2	0.2
	M	MSC	13.3	1.0	0.9
	M	Pericytes	14.9	0.4	0.4
Changepoint 1 (ψ_1)	F	HBSS	12.1	0.3	0.2
	F	MSC	15.5	0.5	0.5
	F	Pericytes	15.2	0.8	0.8
Changepoint 2 (ψ_2)	M	HBSS	16.9	0.3	0.3
	M	MSC	15.5	0.7	0.7
	M	Pericytes	17.6	0.5	0.4
Changepoint 2 (ψ_2)	F	HBSS	19.6	1.1	1.0
	F	MSC	18.0	0.4	0.4
	F	Pericytes	17.6	0.6	0.6

Table 1: Estimates and standard errors for the parameters of model (1) obtained via fitting the approximation (2) along with robust counterpart of the standard errors

Comparison	Changepoint		
	χ^2	df	p-value
Sex within HBSS	23.61	1	< 0.001
Sex within MSC	3.53	1	0.060
Sex within Pericytes	0.14	1	0.713
Pericytes = MSC(F)	0.75	1	0.688
Pericytes + MSC(F) = HBSS(M)	10.16	1	0.001
MSC(M) = HBSS(M)	0.25	1	0.615
MSC(M) = HBSS(F)	1.35	1	0.245

Table 2: Comparisons of changepoints (ψ_1)

Figure 2: Predicted subject specific response curves (MSC males)



Acknowledgements

This research received partial financial support from Conselho Nacional de Desenvolvimento Cient´ıfico e Tecnol´ogico (CNPq, grant 3304126/2015-2) and Fundac¸˜ao de Amparo `a Pesquisa do Estado de S˜ao Paulo (FAPESP, grant 2013/21728-2), Brazil.

References

1. Coatti, G.C. (2015). Avalia¸c˜ao do potencial terapˆeutico de pericitos e de c´elulas mesenquimais no camundongo SOD1, modelo animal para esclerose lateral amiotr´ofica. *Tese de doutorado, Departamento de Bioc^encias, Universidade de S˜ao Paulo*. http://www.teses.usp.br/teses/disponiveis/41/41131/tde-14012016-143346/pt-br.php
2. Fasola, S., Muggeo, V.M.R. and Ku¨chenhoff, H. (2018). A heuristic, iterative algorithm for change-point detection in abrupt change models. *Compuational Statistics* 33, 997-1015.
3. Muggeo, V.M.R., Atkins, D.C., Gallop, R.J. and Dimidjian, S. (2014). Segmented mixed models with random changepoints: a maximum likelihood approach with application to treatment for depression study. *Statistical Modelling* 14, 293-313.
4. Singer, J.M., Rocha, F.M.M. and Nobre, J.S. (2017). Graphical tools for detecting departures from linear mixed models assumptions and some remedial measures. *International Statistical Review* 85, 290-324.