Amyotrophic Lateral Scleroris Functional Score

a Bayesian Analysis

Blerta Begu, Daniele Ceccarelli, Massimiliano Riva

Report
Bayesian Statistics
Professor: Alessandra Guglielmi

Mathematical engineering
February 19, 2020

Abstract

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease that causes the collapse of the nerve cells in the brain and in the spinal cord these are responsible for the trasmission of movement-related impulses hence all patients have to face death within a rage of years that goes from 3 to 5 starting from diagnosis. Only about 25% of patients exceed this threshold. PRO-ACT is an open source dataset, the largest available, that collects different parameters regarding thousands of patients which suffer from this burden. Symptom severity is frequently assessed using a functional scale: ALSFRS. In this project we introduce different models via bayesian tools and by comparing them we choose the best one in order to predict the evolution of the FRS hence of the disease in incoming patients.

Contents

Available dataset

1.1 PRO-ACT

PRO-ACT includes information from over 8500 ALS patients from different clinical trials subdivided in different files. Datas are de-identified to protect patient privacy and different types of information may be available for different patients since multiple trials were merged. During the hospital stay some of the patients received placebo treatments, while others received experimental treatments, however the medications tested in these specific trials were found to be even worst than placebo with respect to their effects on ALS progression.

For every type of information available there is a data file. Each subject is identified by an ID and the specific assessment is identified by a record (each subject has multiple records). The assessments are separated into different files according to the type. For each dataset, the time at which an assessment was taken (a record was created) is listed as the assessment's delta. Delta is given as days from the trial onset which is considered as the null time. A negative delta lists events occurring before hospitalization.

1.2 The Functional Rating Score

The available infromation are wide hence for the aim of this project we focus on this important variable.
The ALSFRS scale is a list of 10 assessments regarding motor functions with each measure ranging from 0 to 4 where 4 is the highest, meaning normal motor function, and 0 meaning total paralysis of the area. The scores

obtained from the individual questions are then summed together to generate a number that constitutes the ALSFRS score.
The questions in the survey concern:

1. speech
2. salivation
3. swallowing
4. handwriting
5. cutting food and handling utensils with or without gastrostomyis
6. dressing and hygiene
7. turning in bed and adjusting bed clothes
8. walking
9. climbing stairs
10. respiratory capacity

ALSFRS-R is a modified version of the ALSFRS. Whereas in the ALSFRS there are 10 assessments, in the ALSFRS-R one of the assessments, the respiratory capacity, was further divided into three questions: Dyspnea, Orthopnea, Respiratory Insufficiency, to better reflect the importance of respiratory changes within the scale. Therefore ALSFRS-R, contains 12 questions (9 of these identical to the traditional ALSFRS) and a maximal score of 48.

1.2.1 Longitudinal Data

This type of model lends itself well since our dataset consist in repeated measures over time indeed longitudinal data, sometimes called panel data, are a collection of repeated observations of the same subjects, taken from a larger population, over some time and they are useful for measuring changes.

Final dataset

Instead of working on all the information available we created a unique dataset which contains the informations that in literature are find to be the more useful in order to make a prediction of the disease progression.
The final dataset contains 6119 patients and 20 variables.

2.1 Variables

The measurements contained in the new data are the following:

- Delta: day from the beginning of the trial in which the observations were recorded, we'll indicate it as t
- ALSFRS-R: patients may have ALSFRS or ALSFRS-R but just few of them have both the scores. We chose to use the second as our response variable since we can compute ALSFRS-R from ALSFRS by triplicating the last score of ALSFRS (respiratory capacity) in such a way we got a larger set of patients.
- Onset Site: the site of disease onset can be a limb (limb onset) or the muscles controlling speaking and swallowing (bulbar onset) or occasionally both.
- Onset delta: the time between disease onset and the first time the patient was tested in a trial.
- FVC: forced vital capacity is the volume of air that can forcibly be blown out after full inspiration, measured in liters
- Age and sex

- BMI-0: body mass index taken at time zero
- BMI-diff: (BMI(t)-BMI(0))/BMI(0)
- SGPT, SGOT, bilirubin: liver-related tests
- Glucose, hematocrit, hemoglobin, RBC, WBC, urinePH
- Riluzole, study treatment: the first is a dummy, 0 if the patient wasn't treated with Riluzole and 1 otherwise; the second says if patients recieved treatments or only placebo.

2.2 Data processing

The problems that we had to face are: the presence of missing values and time incoherence in time-varing variables. In fact it happens that in a same patient observations come out to be recorded in decoupled deltas, in order to have only one variable Delta for each patient we interpolated the faulty variables, our reference Deltas were the ones corresponding to the FRS that for us will be our response in the models, that means that functional rating scores are never interpolated. Whereas for non time-varing variables ad the body mass index at time zero and the variable age we replaced the large presence of missing values with the corresponding gender mode. For all other missing values we used MICE package.

2.2.1 MICE algorithm

R package mice imputes incomplete multivariate data by chained equations. Let Y_j with $j = 1, \ldots, p$ be one of p incomplete variables, where $Y = (Y_1, \ldots, Y_p)$. The observed and missing parts of Y_j are denoted by Y_j^{obs} and Y_j^{mis}, respectively, so $Y^{obs} = (Y_1^{obs}, \ldots, Y_p^{obs})$ and $Y^{mis} = (Y_1^{mis}, \ldots, Y_p^{mis})$ stand for the observed and missing data in Y. The number of imputation is equal to $m \geq 1$. The h-th imputed data set is denoted as $Y^{(h)}$ where $h = 1, \ldots, m$. Let $Y_{-j} = (Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p)$ denote the collection of the $p-1$ variables in Y except Y_j.

Let the hypothetically complete data Y be a partially observed random sample from the p-variate distribution $\mathcal{L}(Y_j|\theta)$. We assume that the multivariate

distribution of Y is completely specified by θ, a vector of unknown parameters. The MICE algorithm obtains the posterior distribution of θ by sampling iteratively from conditional distributions of the form:

$$\begin{gathered}\mathcal{L}(Y_1|Y_{-1},\theta_1)\\ \vdots\\ \mathcal{L}(Y_p|Y_{-p},\theta_p)\end{gathered}$$

The parameters $\theta_1,\ldots,\theta_p$ are specific to the respective conditional densities. Starting from a simple draw from observed marginal distributions, the t-th iteration of chained equations is a Gibbs sampler that draws

$$\begin{aligned}\theta_1^{*(t)} &\sim \mathcal{L}(\theta_1|Y_1^{obs},Y_2^{(t-1)},\ldots,Y_p^{(t-1)})\\ Y_1^{*(t)} &\sim \mathcal{L}(Y_1|Y_1^{obs},Y_2^{(t-1)},\ldots,Y_p^{(t-1)},\theta_1^{*(t)})\\ &\vdots\\ \theta_p^{*(t)} &\sim \mathcal{L}(\theta_p|Y_p^{obs},Y_2^{(t)},\ldots,Y_{p-1}^{(t)})\\ Y_p^{*(t)} &\sim \mathcal{L}(Y_p|Y_p^{obs},Y_2^{(t)},\ldots,Y_p^{(t)},\theta_p^{*(t)})\end{aligned}$$

where $Y_j^{(t)} = (Y_j^{obs},Y_j^{*(t)})$ is the j-th imputed variable at iteration t. We also observe that previous imputations $Y_j^{*(t-1)}$ only enter $Y_j^{*(t)}$ through its relation with other variables, and not directly. Convergence can therefore be quite fast.

Mixed Effect Models

We suspect the functional rating score to be strongly connected with the behaviour of other variables. In order to investigate our a prior beliefs we will work in the mixed effect framework. This contest is acceptable since we have repeated measures for each ID and each patient will be then considered as a self specific group which has its own trend but of course also between groups interactions should be taken into account since the presence of the disease is a binding factor. In this section the theoretical setting is exposed.

3.1 Definition

Model 3.1.1. *(Linear Mixed effects)*

$$Y_{ij} = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta}_j + \varepsilon_{ij} \quad \forall j = 1, \dots, m \ \ \forall i = 1, \dots, n_j$$

where j is the group index. i is the group specific observation and n_j is the number of observation in group j.
Equivalently it can be written as:

$$\begin{aligned}
\boldsymbol{Y}_j &\overset{iid}{\sim} \mathcal{N}_{n_j}\left(X_j, \sigma^2 I_{n_j}\right) \\
\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m | \boldsymbol{\theta}, \Sigma &\overset{iid}{\sim} \textit{multivariate-}\mathcal{N}\textit{ormal}(\boldsymbol{\theta}, \Sigma) \\
\varepsilon_{ij} | \sigma^2 &\overset{iid}{\sim} \mathcal{N}(o, \sigma^2) \\
\boldsymbol{\theta}, \Sigma &\sim \Pi(\boldsymbol{\theta}, \Sigma) = \Pi(\boldsymbol{\theta}) \times \Pi(\Sigma) \\
\sigma^2 &\sim \Pi(\sigma^2)
\end{aligned}$$

There is also an alternative parametrization of the model which underlines better the presence of fixed and random effects:

Model 3.1.2. *(Linear Mixed effects)*
By setting $\boldsymbol{\beta}_j = \boldsymbol{\theta} + \boldsymbol{\gamma}_j \quad \forall j = 1, \dots, m$ *the single response can be modeled as:*

$$Y_{ij} = \boldsymbol{x}_{ij}^\top \boldsymbol{\theta}_j + \boldsymbol{z}_{ij}^\top \boldsymbol{\gamma}_j + \varepsilon_{ij} \quad \forall j = 1, \dots, m \ \forall i = 1, \dots, n_j$$

where $\boldsymbol{\theta}_j$ *is the so called fixed effect that is constant across groups and* $\boldsymbol{\gamma}_j$ *is the random effect which is group specific.* $\boldsymbol{x}_{ij}$ *and* $\boldsymbol{z}_{ij}$ *could be vectors of different lengths that may or may not contain overlapping covariates.*
Usually priors are set as follows:

$$\begin{aligned}
\boldsymbol{\gamma}_j | \Sigma &\overset{iid}{\sim} \textit{multivariate-}\mathcal{N}\textit{ormal}(\mathbf{0}, \Sigma) \quad \forall j = 1, \dots, m \\
\varepsilon_{ij} | \sigma^2 &\overset{iid}{\sim} \mathcal{N}(o, \sigma^2) \\
\boldsymbol{\theta}, \Sigma, \sigma^2 &\sim \Pi(\boldsymbol{\theta}, \Sigma, \sigma^2) = \Pi(\boldsymbol{\theta}) \times \Pi(\Sigma) \times \Pi(\sigma^2) \\
\boldsymbol{\theta} &\sim \textit{multivariate-}\mathcal{N}\textit{ormal}(\boldsymbol{\mu}_0, L_0) \\
\Sigma &\sim \textit{inverse-Wishard}(S_0^{-1}, \eta_0) \\
\sigma^2 &\sim \textit{inverse-gamma}\big(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\big)
\end{aligned}$$

Throughout our project we will make use of the predictive distributions to check wether or not the models give a good approximation. In this contests they are defined as follows:

Predictive distribution 3.1.3. *(New sumbject from a new group)*

$$\mathcal{L}\big(\boldsymbol{Y}_{m+1}^{new}, \boldsymbol{\beta}_{m+1} | data, X^{new}\big) = \int \mathcal{L}\big(\boldsymbol{Y}_{m+1}^{new} | \boldsymbol{\beta}_{m+1}, X^{new}\big) \mathcal{L}\big(\boldsymbol{\beta}_{m+1} | \boldsymbol{\theta}, \Sigma\big) \mathcal{L}\big(d\boldsymbol{\theta}, d\Sigma, d\boldsymbol{\beta}_{1:m} | data\big)$$

Predictive distribution 3.1.4. *(New sumbject from an existing group)*

$$\mathcal{L}\big(\boldsymbol{Y}_j^{new} | data, X^{new}\big) = \int \mathcal{L}\big(\boldsymbol{Y}_j^{new} | \boldsymbol{\beta}_j, X^{new}\big) \mathcal{L}\big(d\boldsymbol{\beta}_j, d\boldsymbol{\theta}, d\Sigma | data\big)$$

Implemented models

From now on our response variable Y_{ij} will be the functional rating score of patient j at observation i. The total number of patients is $m = 6119$. We refer always to the theoretical setting introduced previously.

4.1 Model evaluation

We decided to split our dataset in 3 parts: a training set and two different test sets. Firstly, we select a group of patients to use them to implement our models "alsfrs-train". Another group of ID is used as test set "alsfrs-test". Then we take our training set and we split it in two parts: for some patients, (50%) of the training set, we take the second half temporal evaluation and we put it in a new test set, and we call it "alsfrs-train-patient-test-observation". We have done this second split for two reasons:

1. we can test our model on two different tests set, with different characteristics
2. mixed effect model make prediction in a different way if the patient ID is present or not in the training set (new group vs existing group) as we'll explain below.

To evaluate our model we compute two different prediction errors, and we evaluate on the three different dataset that we have alsfrs-train, alsfrs-train-patient-test-observation, alsfrs-test. The first error is a mean absolute error:

$$err_{1_{ij}} = \left| \mathbb{E}\left[Y_{ij}^{sim} \right] - Y_{ij}^{true} \right| \approx \left| \left(\frac{1}{M} \sum_{k=1}^{M} Y_{ij}^{(k)} \right) - Y_{ij}^{true} \right|$$

where we used the MCMC approximation and M is the number of iterations.

$$err_1 = \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n_j} err_{1_{ij}} \quad \text{with } n = \sum_{j=1}^{m} n_j$$

The second error takes in consideration the variability of the prediction:

$$err_{2_{ij}} = \left|\left(\frac{1}{M}\sum_{k=1}^{M} Y_{ij}^{(k)}\right) - Y_{ij}^{true}\right| + |CI_{95\%}|_{ij}$$
$$err_2 = \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n_j} err_{2_{ij}}$$

4.2 Dealing with the time

In order to assess the goodness of the Longitudinal approach we begin by introducing as an only variable the time t, which compares in the dataset with the name *Delta*.

4.2.1 Longitudinal model

We start with the basic linear model that reads:

$$Y_{ij} = \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}_{1j} t_{ij} + \varepsilon_{i,j} \qquad \forall j = 1, \dots, m \ \ \forall i = 1, \dots, n_j$$

where $\boldsymbol{\beta}_{0j} = \theta_0 + \gamma_{0j}$ and $\boldsymbol{\beta}_{1j} = \theta_1 + \gamma_{1j}$ hence we consider both fixed and random intercepts and slopes. And as written in the general setting $\varepsilon_{ij}|\sigma^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.
Since the motor efficiency for every patient decreases in time we expect to have $\theta_1 < 0$ (a negative fixed effect slope).
The model has been implemented using Stan with imputs: $iter = 10000, thin = 5$ and $warmup = 1000$ and priors:

$$\begin{aligned} \theta_0, \theta_1 &\overset{iid}{\sim} \mathcal{N}(0,1) \\ \gamma_{0j}|\sigma_0^2 &\sim \mathcal{N}(0, \sigma_0^2) \qquad \forall j = 1, \dots, m \\ \gamma_{1j}|\sigma_1^2 &\sim \mathcal{N}(0, \sigma_1^2) \qquad \forall j = 1, \dots, m \\ \sigma^2, \sigma_0^2, \sigma_1^2 &\overset{iid}{\sim} \text{inv-gamma}(2,1) \end{aligned} \tag{4.1}$$

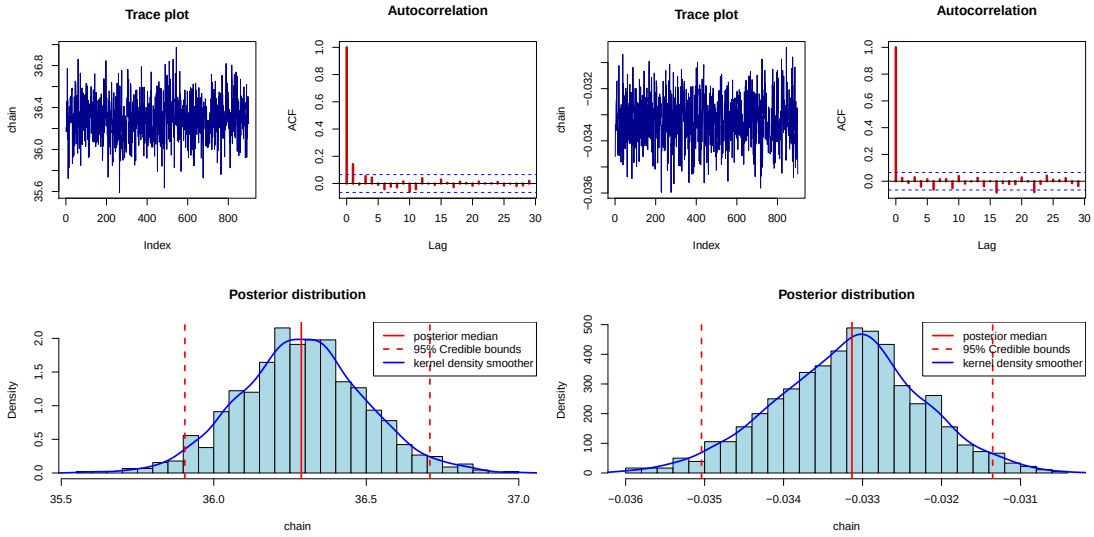


Figure 4.1: On the left refering to θ_0 and on the right refering to θ_1

As an outcome we have the following plots: As axpected the 95% posterior credible interval of θ_1 is all negative. In both cases we can deduce that the time is a needful covariate for our goal, but we suspect that a linear dependence on the time might not be sufficient for this reason follows a model which considers basis functions.

4.2.2 Cosine parametric model

By looking at the partial autocorrelations of the functional rating scores for each patient one can clearly see a wavy dependence for this reason we implemented the following basis expansion:

$$Y_{ij} \overset{iid}{\sim} \mathcal{N}(f(t_{ij}), \sigma^2)$$

where:

$$f(t_{ij}) = \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}_{1j} t_{ij} + \sum_{k=1}^{3} \boldsymbol{\alpha}_{kj} \cos\Big(\frac{k\pi(t_{ij} - t_{(1)j})}{t_{(n_j)j} - t_{(1)j}}\Big) \tag{4.2}$$

with $t_{(1)j}$ being the minimum *Delta* of subject j and $t_{(n_j)j}$ the maximum, moreover $\boldsymbol{\alpha}_{kj}$ has also both fixed and random parameters. The model has been implemented using Stan with imputs: $iter = 10000, thin = 5$ and

$warmup = 1000$ and priors choosen proceeding similarly as in (4.1) for both fixed and random effects.
Posterior credible intervals for the fixed parameters of $\boldsymbol{\alpha}_{kj}$:

	2.5%	50%	97.5%
α_1	1.514778	1.821671	2.157200
α_2	-0.18513127	-0.12183194	-0.06732484
α_3	0.1734543	0.2491597	0.3305144

Since the null point is not covered in none of the intervals it means that these basis functions are usefull for modeling our response.

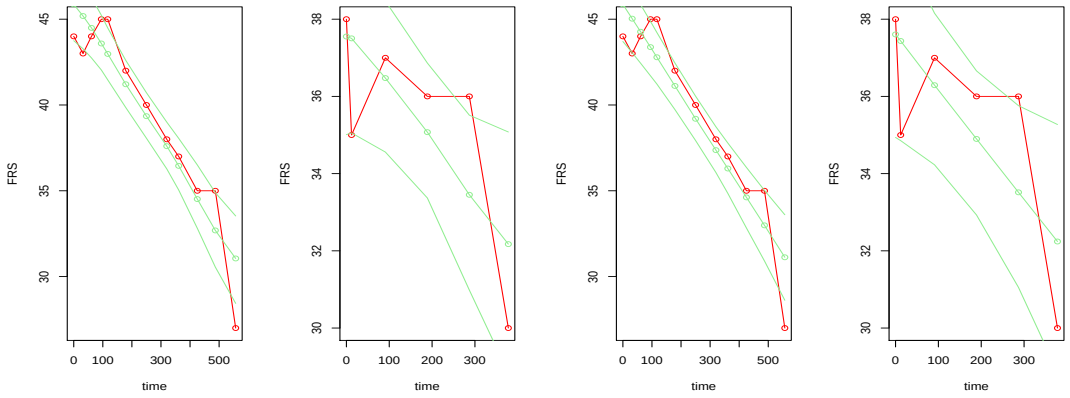


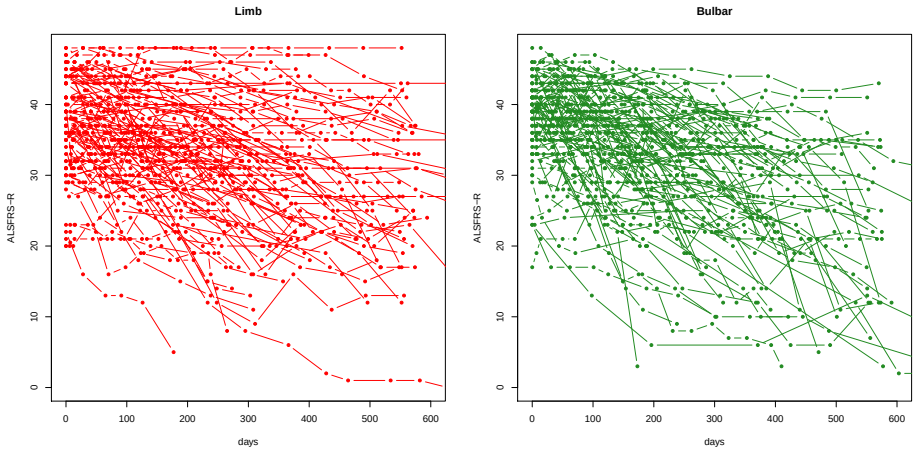
Figure 4.2: On the left 95%CI for longitudinal model and on the right for cosine model. Bounds in the cosine model are slightly smoothed.

4.3 Multicovariate models

We firstly add to our model only one covariate: *Onset site*, follows the investigation of the treatment variables and then before introducing the multicovariate model we implement a variable selection using horseshoe priors.

4.3.1 Onset site

We introduce the variable Onset site to the longitudinal model. This comes from looking at the trend of the functional scale in the two different groups:



As one can deduce from the plots bulbar ALS tends to be more acute than limb onset type. The model in this case is the one that follows:

$$Y_{ij} = \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}_{1j} t_{ij} + \theta_2 d_j + \theta_3 d_j t + \varepsilon_{ij} \qquad \forall j = 1, \dots, m \;\; \forall i = 1, \dots, n_j$$

where $\boldsymbol{\beta}_{0j} = \theta_0 + \gamma_{0j}$ and $\boldsymbol{\beta}_{1j} = \theta_1 + \gamma_{1j}$ hence we consider both fixed and random intercepts and slopes for the time variable and only fixed for the dummy variable d_j which represents the onset site for patient j $d_j = 1$ in the case of Limb type ALS and fixed variable for the interaction between the dummy and the time. And as written in the general setting $\varepsilon_{ij}|\sigma^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.
The model has been implemented using Stan with imputs: $iter = 10000, thin = 5$ and $warmup = 1000$ and priors choosen proceeding similarly as in (4.1) for both fixed and random effects.
The posterior 95%CI of the parameter of interest θ_3 is:

	2.5%	50%	97.5%
θ_3	0.003590462	0.009774519	0.015824146

The zero is not covered by the interval therefore we deduce onset site to be a good variable to take into consideration, moreover the interval is all positive so it convalidates the behaviour represented in the introductory figure.

	Training set	Train with new obs.	Test set
err_1	1.353407	3.808556	6.555238
err_2	10.62259	20.34279	40.21576

4.3.2 Treatments

Now we add to the previous model two other dummy variables one for the Riluzole drug and one for the *study treatment* variable leading to the model:

$$Y_{ij} = \boldsymbol{\beta}_{0j} + \boldsymbol{\beta}_{1j}t_{ij} + \theta_2 d_{1j} + \theta_3 d_{1j}t + \theta_4 d_{2_j} + \theta_5 d_{2j}t + \theta_6 d_{3j} + \theta_7 d_{3j}t + \varepsilon_{ij}$$
$$\forall j = 1, \dots, m \;\; \forall i = 1, \dots, n_j$$

where d_{1j}, d_{2j}, d_{3j} represent the dummies of *onset-site, Riluzole, study treatment* for patient j. It is currently known that there aren't medications that give good feedbacks in treating the disease hence we expect the new dummies to be unusefull. The model has been implemented using Stan with imputs: $iter = 10000, thin = 5$ and $warmup = 1000$ and priors choosen proceeding similarly as in (4.1) for both fixed and random effects.
The posterior 95%CI of the parameters of interest θ_5 and θ_7 are:

	2.5%	50%	97.5%
θ_5	-0.018107029	-0.003197521	0.012007774
θ_7	-0.008408403	-0.003589222	0.001239734

Both contain the zero point so they are not considered to be relevant variables.

	Training set	Train with new obs.	Test set
err_1	1.353605	3.7856	6.587278
err_2	10.61728	20.29393	40.2572

4.3.3 Variabel selection - final model

After these simple models, we decided to add all the covariates that we have collected through our data preprocessing.
In addiction to information about the type of ALS (Bulbar or Limb, contained in the column "Onset_site") and Medication (Treatment and Riluzole), we can use different information from each patient to model its functional score, for example:

- Personal Information: Age and Sex;
- BMI:
  - Initial BMI at time zero (fixed)
  - Difference in % from the initial BMI (longitudinal)
- Onset Delta (how many days before Time 0 the patient had the first symptoms);
- Lab data: FVC, SGPT, SGOT, HEMATOCRIT, HEMOGLOBIN, RBC and WBC.

Variable selection

With so many covariates used in the fixed effect part of the model, in addiction to the random effect part (two coefficients for each patient if we use just intercept and slopes for Delta), we have to face an overfitting problem.
For this reason, we use a regularization in the prior of theta, in particular a Horseshoe prior (Carvalho, Polson, and Scott 2009).

$$\begin{aligned} \theta_p \,|\, \tau,\, \lambda_p &\overset{iid}{\sim} N(0, \tau\lambda_p) \\ \lambda_p &\overset{iid}{\sim} C^+(0,1) \\ \tau &\sim C^+(0,\tau_0) \\ \forall p &= 1, \ldots, P \end{aligned}$$

If we want to understand how this regularization works, we can see the comparison of the unit ball from classical regularization (for example Ridge, Lassu, Cauchy) and the Horseshoe.

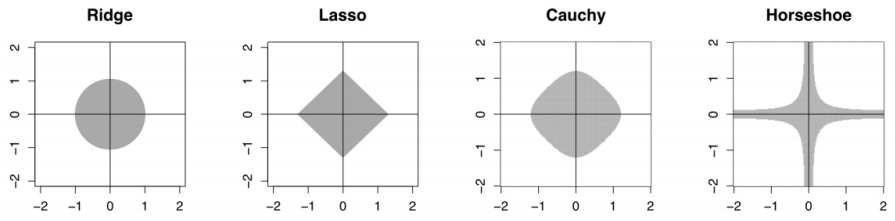


Figure 4.3: Regularizations

We can see that, compared to the first three, the unit ball in horseshoe framework includes all the axes: for this reason, the prior gives a lot of density to zero, as we can visualize in this image:

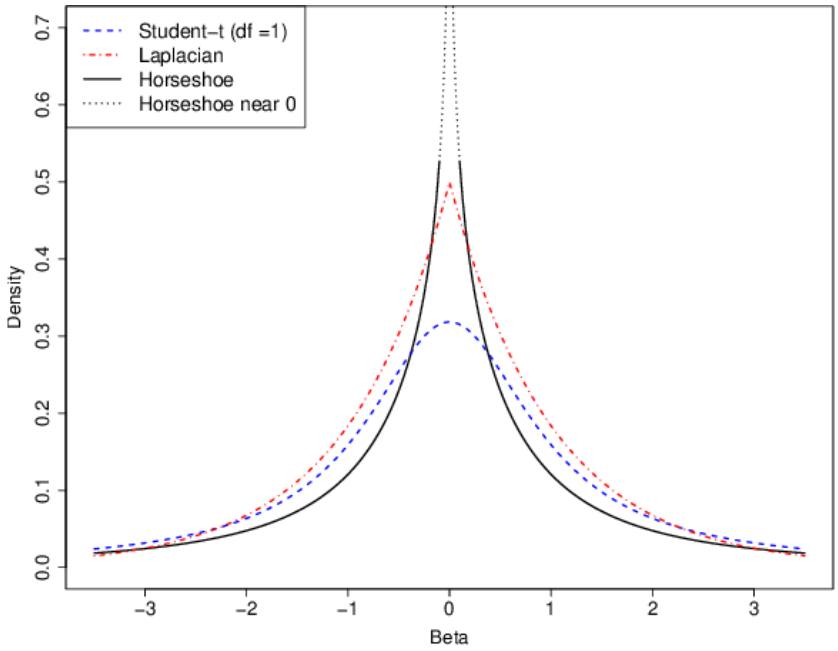


Figure 4.4: Horseshoe density

This regularization help us to understand which covariates are really significative for our model in fact we can discard the thetas that has all the density concentrates around zero.
To run the model in BRM (or Stan) with the horseshoe prior, we have to change some control parameters of our MCMC to avoid divergent transitions, in particular:

- Increase Adapt Delta above the standard value 0.8 (we set it to 0.9);
- Increase max treedepth above 10 (we set it to 15)

and of course these changes increase a lot the computational time.

As we can see in these plot, there are some covariates that we can erase from our model:

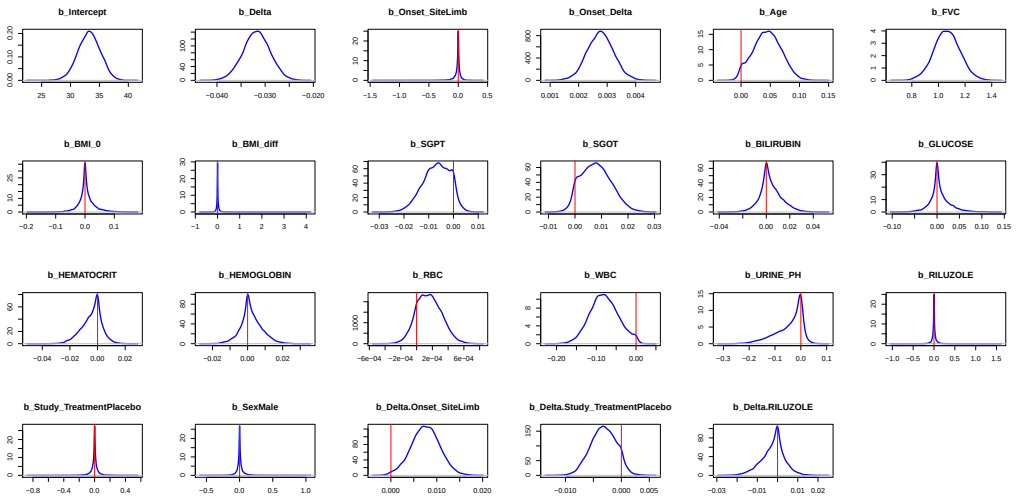


Figure 4.5: Density of Theta with Horseshoe prior

After this visual analysis, we choose to discard BMI_0, BMI_{diff}, Bilirubin, Glucose, Hematocrit, Hemoglobin, Urin PH, Riluzole and Sex.

Final model

For the final model, we add to the cosine parametric model previously introduced the covariates that we kept after the variable selection.

$$Y_{ij} = f(t_{ij}) + \boldsymbol{x}_{ij}^{\top}\boldsymbol{\theta}_j^* + \varepsilon_{ij} \quad \forall j = 1, \dots, m \;\; \forall i = 1, \dots, n_j$$

Where $f(\cdot)$ was defined in (4.2), $\boldsymbol{x}_{ij}$ is the vector containing all the observations at time i of patient j of the other covariates with iterations included and $\boldsymbol{\theta}_j^* = (\theta_2, \dots, \theta_{11})$. The results are consistent with the previous models, and we achieve better result in terms of error.

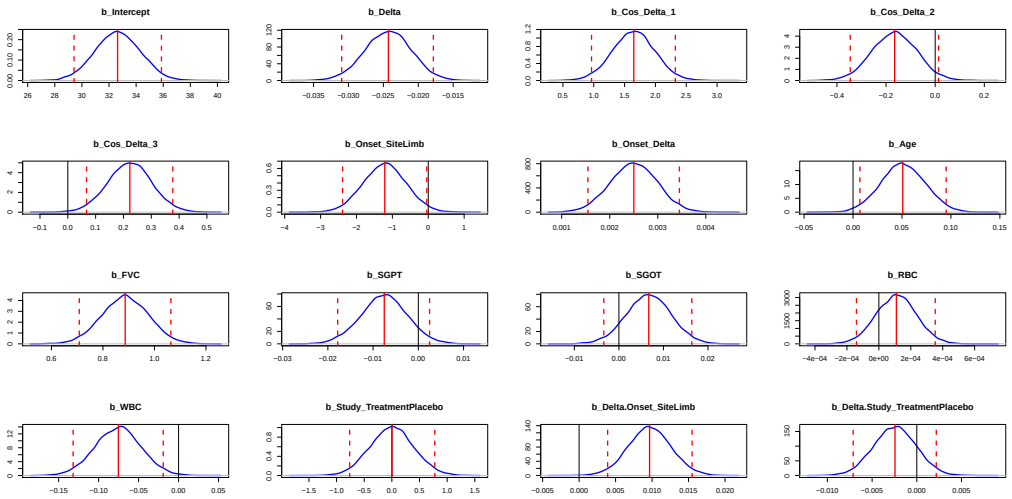


Figure 4.6

For what concernes the coefficients estimate of the model (now with Normal priors as in the first models), we can see that they are all statistically significative except from some lab data (like SGPT, SGOT and RBC) and Study Treatment, in accordance with what we found in the "Treatment Model". While the coefficient of "Study Treatment" is perfectly centered in zero, the coefficient of Interaction between Study Treatment and Delta contains zero in its $CI_{95\%}$ but it is close to the boundary lines, for this reason we left Study-Treatment in our model.

	Training set	Train with new obs.	Test set
err_1	1.048792	3.582663	5.977326
err_2	9.024221	20.45746	35.94814

If we focus on the errors for the second case, where we have the patients in the training set but we evaluate the score on some new temporal observations (that is what a doctor should do in a real case: monitoring a patient for a few months and try to predict how he can behave in the future), we have an err_1 of 3.5, that is quite good on a score that ranges from 0 to 48, and an err_2 of 20, that means an uncertainty of plus or minus 8 points more or less (in the case where we consider a credible interval of 95 %).
The error on the training set is very small (err_1 is 1.04) because we can predict well the behaviour of these patients using all the covariates in the fixed effect part but also using the random effect part, that can be used to model the variability among the patients.
Finally, the error on some new patients, i.e. patients such that their ID is not present in the training dataset, is bigger than the previous two cases, and this is due to the fact that now we can't sample directly from the random effect specifically of these patients, because we don't have it in our model, but we have to sample it from the posterior distribution of the random effects that is basically a normal distribution with zero mean and a big variance (thanks to the eteronegeneity of our training patients).
We will see some examples of prediction of train patient but on new temporal observation (that is, for the reason explained above, the most intresting prediction we can make) in the last section of this report.

Model checking

We recall that we made nested splittings in our dataset, first we divided it in two parts: train and test and then we took the train data and for the first half of patients we put their second half of time observation in a new dataset that was named *train-patient-test-observation*.

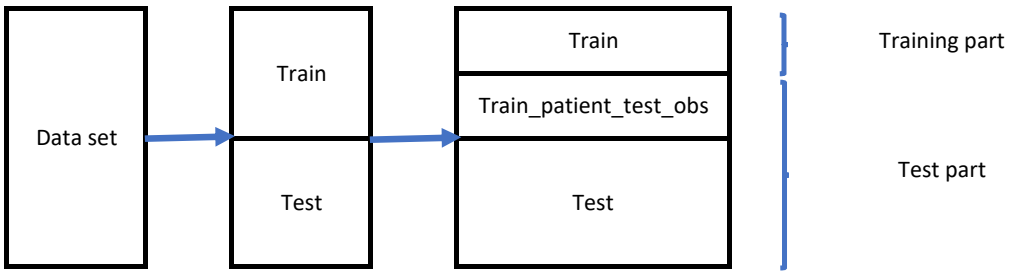


Figure 5.1: Data subdivision

5.1 Prediction plots

Firstly we check the final model using Train dataset. It can be seen below a very good prediction, thanks to the fact that we we can sample directly from the random effect of a patients that is into the training set. In the plot the

blue trend represents the true values of the specific ID.

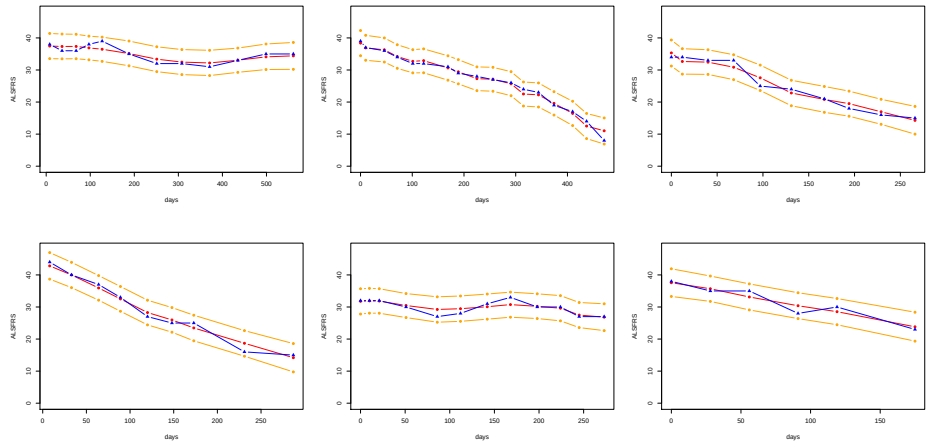


Figure 5.2: Model checking: $CI_{95\%}$

Since the checking gave good feedbacks we went on by using *train-patient-test-obs*, also here we expect nice results since we can sample directly from the random effect of a patient in the training part but $CI_{95\%}$ expand in time since we are predicting new temporal observations.

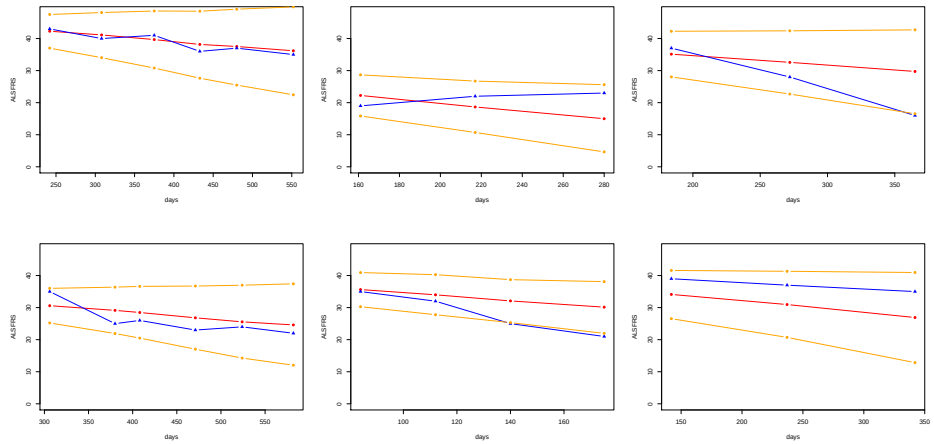


Figure 5.3

This type of prediction is the most interesting in the real applications since we expect doctors to monitor patients for a few months and use the observations to make suppositions about the future decay of the disease. The last case that we comment is the prediction of a new patient that doesn't belong to the training part (Test set).

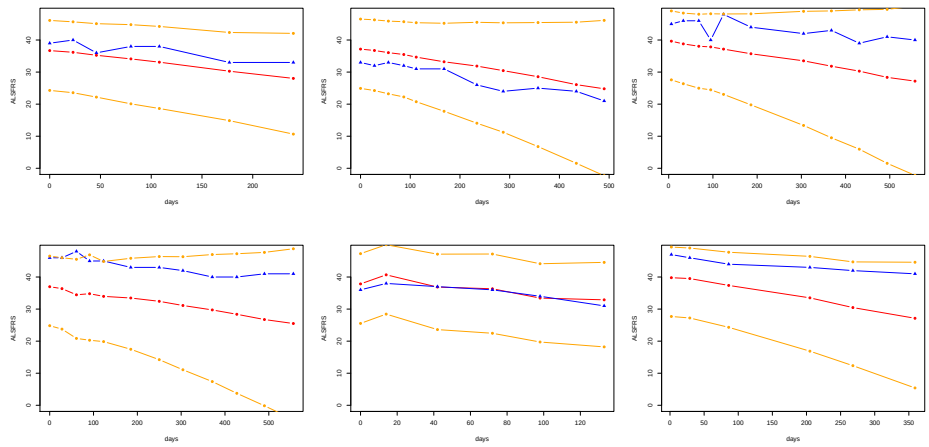


Figure 5.4

In this case as we expected we found wider bands than before; this is due to the fact that we are sampling from the posterior of the random effect that behaves as a normal distribution with zero mean and a large variance since the patients in our training set are very eterogeneous and so the model uses big random effects in the train prediction.

As commented in the second case we focused in predicting patients in *train-patient-test-observation*. Follows two examples:

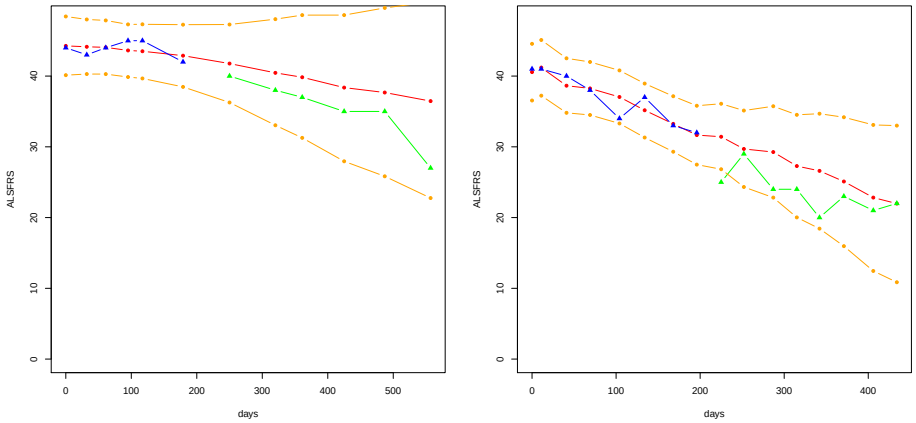


Figure 5.5

We can see in blue the trends of the patient in the train dataset while in green the trend of the patient in *train-patient-test-observation* where we wanted to make prediction, we expected the bands to widen up after the blue part but predictions are still good.

5.2 Variability among patients

This model is usefull not only to make prediction but also to analyze the common features of the patients in the training set. In particular, the random effects tells us something about the variability among patients. We plotted the mean of each random's effect posterior distrubution and of course all the points are located around the zero. We have as an example below the plot of intercept vs slope in which we see a different partition of the space, a good number of points is concentred in the right side but on the left side we have a larger variability.

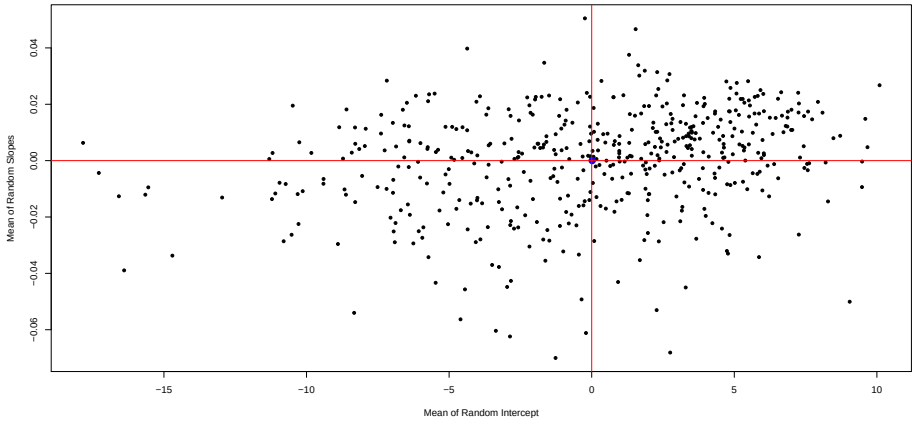


Figure 5.6

It is interesting to understand why some patients are so far from the null point, for this reason we selected some of them:

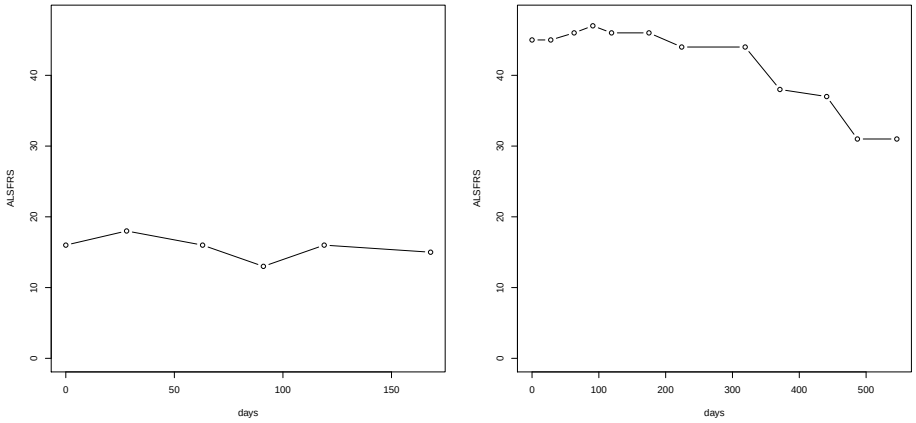


Figure 5.7

Above we can see a kind of patients which entered in the trial at an advanced stage of the disease (on the left of the plot) and at a beginning stage. The fact that there is a large variability in the intercepts means that in our model there isn't a covariate which is able to fully explain this behaviour. In

the model we have Onset delta (the time between disease onset and the first time the patient was tested in a trial) but it is not suffient in this way and so the model needs random effects to adjust the trends.
For what concernes the mean slopes distribution, we plot this two different patients:

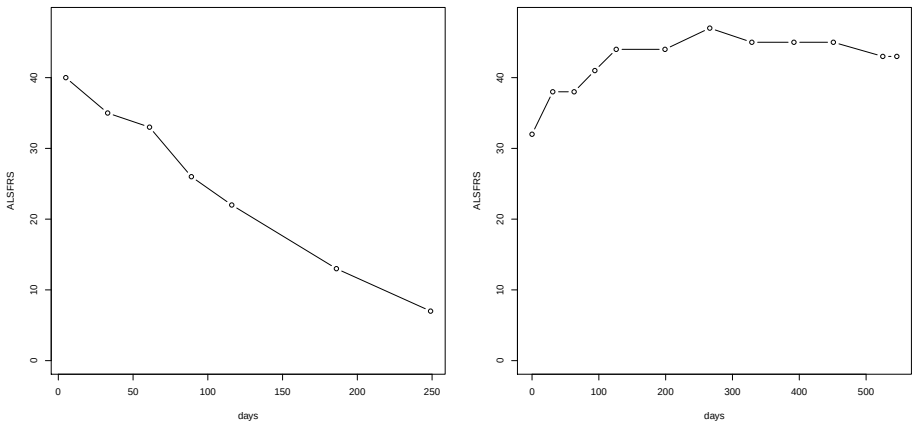


Figure 5.8

On the left a patient with a very fast decreasing, but is more interesting the patient on the right which has a mean random slopes bigger than zero because as we can see, he has a postive trend, that is quite abnormal for a disease like ALS that is neurodegenerative. This patient is a *slow progression type*, where the he or she stays alive for a longer time than usual; another explanation of this trend could be also a human error on the calculation of the FRS.

And finally a comment on the random effect for the cosine basis, for example the cos_1. As we can see below, we have on the left a patient that has first a fast decrease then a period of stabilization and again a new decrease; on the other side, we have a patient with a first part of no decrease, than a rapidly decrease and finally a stabilization part. Both these behaviour validate the use of cosine part also in the random effect because not all the patients have these weavy trends, some of them have a linear one.

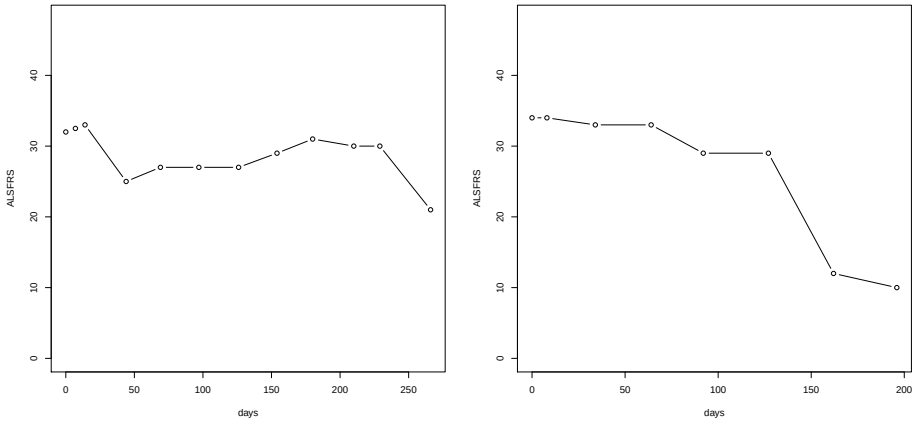


Figure 5.9

References

[1] Paul-Christian Bürkner; Advanced Bayesian Multilevel Modeling with the R Package brms *The R Journal*

[2] Stef van Buuren, Karin Groothuis-Oudshoorn; mice: Multivariate Imputation by Chained Equations in R *Journal of Statistical Sotfware* 2011

[3] Carlos M. Carvalho, Nicholas G. Polson, James G. Scott; Handling Sparsity via the Horseshoe

[4] Fernando A. Quintana, Wesley O. Johnson, Elaine Waetjen, and Ellen Gold; Bayesian Nonparametric Longitudinal Data Analysis *Author manuscript* 2017

[5] Robert Kueffner, Neta Zach, [...]Gustavo Stolovitzky; Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression *Computational biology* 2015

[6] Robert Kueffner, Neta Zach, [...]Gustavo Stolovitzky; Stratification of amyotrophic lateral sclerosis patients: a crowdsourcing approach *Nature* 2019

[7] Renate Meyer; An Introduction to Bayesian Nonparametrics *Department of Statistics, University of Auckland, New Zealand*

[8] The PRO-ACT database. https://nctu.partners.org/proact