

2022 BDA II PART 2 KU LEUVEN
MASTER OF STATISTICS & DATA SCIENCE

Assignment – Part 2

Authors:

Luka BEVERIN (luka.beverin@student.kuleuven.be)
Nicolas COZZI (nicolas.cozzi@student.kuleuven.be)
Marie Analiz April LIMPOCO (marieanalizapril.limpoco@student.kuleuven.be)
Khachatur PAPIKYAN (khachatur.papikyan@student.kuleuven.be)

January 27, 2022

Contents

1 Analysis I	1
2 Analysis II	10
3 Analysis III	13
4 Conclusion	16

1 Analysis I

Fit a linear mixed model for SOFA with covariates age and day.
Check the functional form of the covariates in the model.

In the mechanical ventilation (MV) study, the Sequential Organ Failure Assessment (SOFA) score (range 0-24), was measured for 139 mechanically ventilated patients from their first day on MV until Intensive Care Unit (ICU) discharge or day 30 after MV initiation, whichever occurred first.

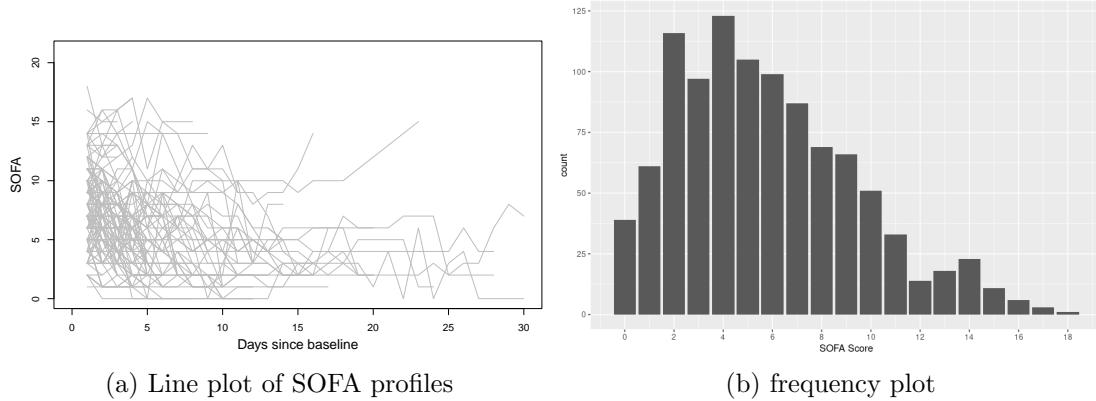


Figure 1: Descriptive plots

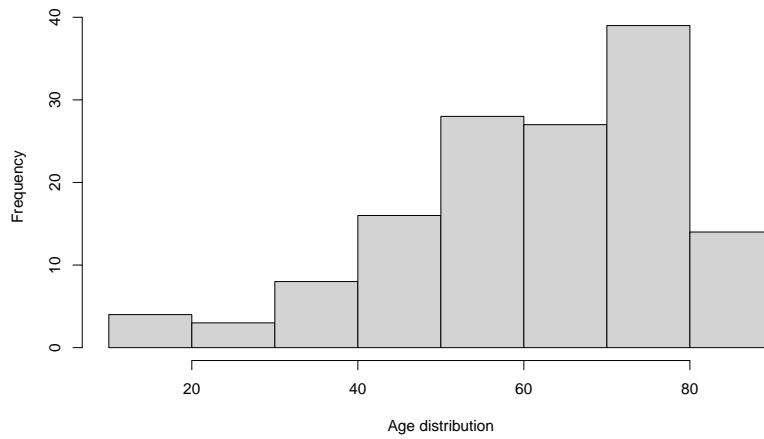


Figure 2: Age distribution of patients

Figure 1a displays the SOFA scores of patients during their stay in ICU plotted against time, which illustrates the unbalanced structure in the data being considered. Meanwhile, Figure 1b shows the positively skewed distribution of SOFA scores. We notice that missing values for SOFA scores were not evenly distributed

amongst the patients, rather they were patient-specific. For example, patient id = 41 had 11 missing values. These missing values were estimated using data augmentation; i.e. sampling the missing data from the conditional distribution and then sampling the parameter from the completed posterior distribution. We also analysed the data when patients with missing values were removed entirely from the data set. Another variable of interest is patient's age, which appears to have a negatively skewed distribution as shown in Figure 2. This means that patients in MV are usually older patients.

We fitted the following linear mixed model (LMM) as a function of age (in years) and time (in days) in the ICU since the study entry:

$$y_{ij} = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{day}_{ij} + b_{0i} + b_{1i} \text{day}_{ij} + \epsilon_{ij}, \quad (1)$$

where y_{ij} is the SOFA score of the i th patient recorded on day j , while b_{0i} is a random intercept and b_{1i} is the random slope assumed to follow $\mathbf{b}_i = (b_{0i}, b_{1i})' \sim N_q(\mathbf{0}, \mathbf{G})$. We first implemented a frequentist linear mixed model using the R package **lme4** wherein we obtained the following maximum likelihood estimates:

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_{b0}^2$	$\hat{\sigma}_{b1}^2$	$\hat{\sigma}_\epsilon^2$
6.93602	0.01746	-0.47298	10.8197	0.1592	2.1719

A random slope model allows each group line to have a different slope, implying that the random slope model allows the explanatory variable to have a different effect for each patient. Implementing the Bayesian framework, we assigned independent vague normal priors $N(0, 10000)$ for the regression coefficients and an inverse Wishart prior for the covariance matrix of the random slope and intercept. We also perform a sensitivity analysis in which we consider Uniform priors $U(0, 1000)$ for the standard deviation of the random effects σ_{b_0} and σ_{b_1} (see [Gelman \[2006\]](#)). Furthermore, we assigned a uniform prior distribution $U(-1, 1)$ to ρ , the pairwise correlation between the random effects. Two chains, each of 100,000 iterations were instantiated with a burn-in rate of 50,000 and a thinning factor of 2 was applied. The **NIMBLE** program was used for this analysis.

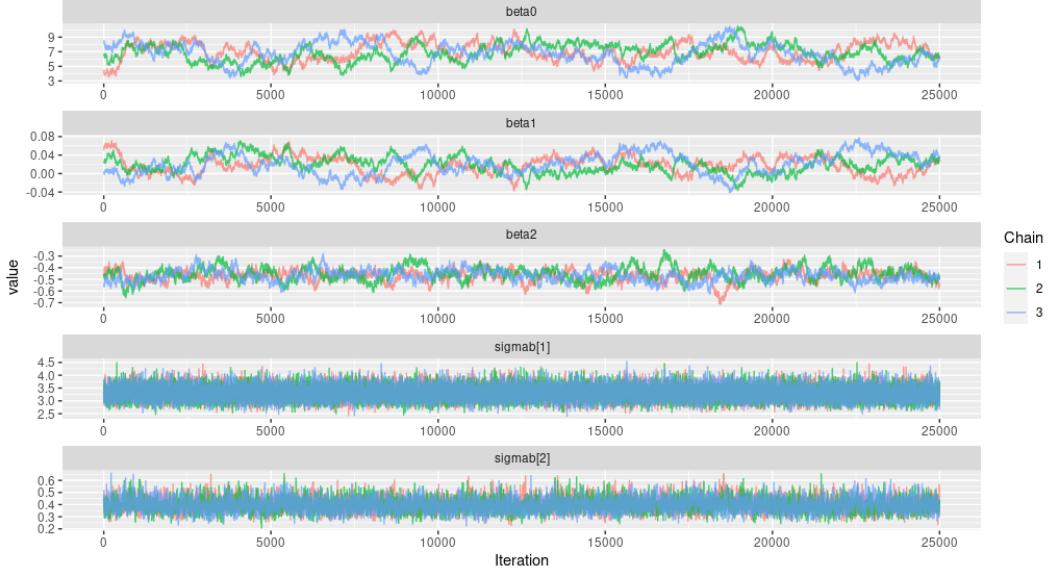


Figure 3: Inverse Wishart prior traceplot

Figure 3 displays the trace plot when the inverse Wishart prior was assigned to \mathbf{G} . There is evidence of marginal mixing as the chains for the fixed effects do not traverse its target distribution quickly. This is also apparent when using the uniform prior and it is likely due to high autocorrelation among the samples, which will be explored later in this section.

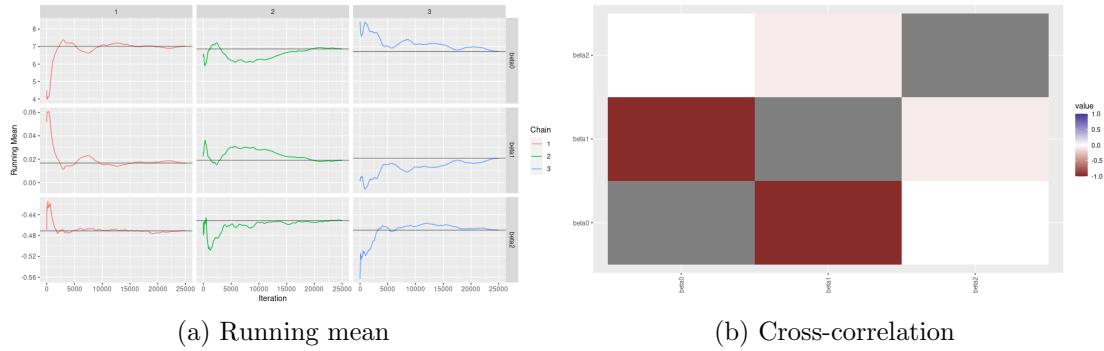


Figure 4: Inverse Wishart prior

Figure 4a displays the running mean of the parameters of interest for the inverse Wishart prior, which allows us to check whether the chain is slowly or quickly approaching its target distribution. The cross-correlation plot displayed in Figure 4b reveals the possible reason for slow convergence of the fixed effects in our model; i.e., a large negative correlation between β_0 and β_1 .

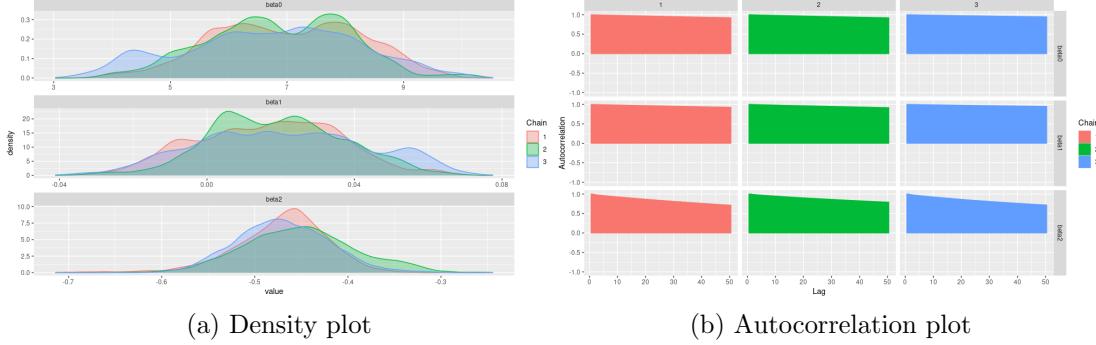


Figure 5: Inverse Wishart prior

A case of consistent high serial correlation is visible in the model, indicating that the parameters in the model may be highly correlated (see Figure 5b). Since the model has high auto-correlation, it requires more iterations to explore the entire posterior distribution. The multi-peak characteristics of the density plots in Figure 5a suggest that the convergence has barely been met when the inverse Wishart prior is assigned to \mathbf{G} . Uniform priors did not substantially improve the results and both models produce similar Watanabe-Akaike Information Criteria (WAIC) values: in the proximity of 3110.

Parameter	Wishart prior	Uniform prior
β_0	6.871	6.953
β_1	0.0189	0.0173
β_2	-0.464	-0.473
σ_{b0}^2	10.76	11.131
σ_{b1}^2	0.155	0.166

Table 1: Mean values

The residuals in Figure 6b tend to stray from the straight line near the tails, indicating that the residuals are likely not normally distributed.

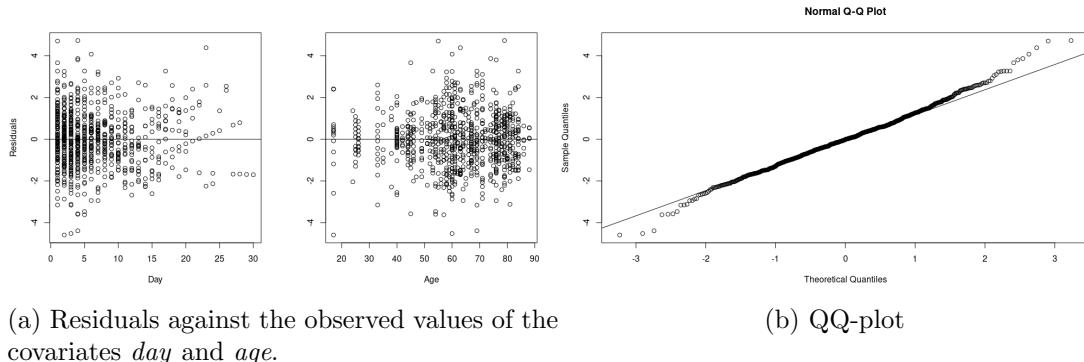


Figure 6: Residual analysis

We also considered a model wherein hierarchical centering was applied; i.e.

$$\log(\text{sofa}_{ij} + 1) = \beta_1 \text{age}_i + b_{0i} + b_{1i} \text{day}_{ij} + \epsilon_{ij}, \quad (2)$$

We added a small positive value 1 before taking the log of sofa to avoid problems related to taking the log of 0. Also, the log transformation was used since sofa is a positive variable. In this model, missing values were estimated using data augmentation from a Bayesian standpoint. The parameters β_0 and β_2 from the first model are now taken to be the mean of the priors for the random slope and intercept respectively. That is, $b_{0i} \sim N(\beta_0, \mathbf{G})$ and $b_{1i} \sim N(\beta_2, \mathbf{G})$ following the same priors as the previous model. After 100,000 iterations with 50,000 burn-in, the trace plots appear stationary except for β_1 as shown in Figure 7. However, the global BGR diagnostics Rhat for all parameters were 1 and the results give a conditional WAIC value of 661.96, a pD or effective number of parameters equal to 148, with posterior estimates shown in Table 2, along with the MLE results, where we see that the Bayesian and frequentist results are very close. Figure 8a displays the posterior predictive distribution of SOFA for this model while Figure 8b shows the posterior predictive check performed wherein the PPC value (0.5004) and the plot of the sum of squares for the observed and predicted values suggest a good model fit. We also attempted to compute the marginal WAIC for this model in Nimble but for 20,000 iterations with 10,000 burn-in, it took 5 hours more or less for 1 run. Hence, in the interest of time, we used conditional WAIC for the subsequent comparisons and also since all models we considered have random effects anyway. The marginal WAIC for this model is 2086.507.

Parameter	Pos mean[2.5%,97.5%]	sd	MC error	Rhat	ESS	MLE
β_0	2.121[2.040,2.203]	0.041	0.0004	1	9787	2.112
β_1	0.068[-0.005,0.141]	0.036	0.0014	1.02	693	0.069
β_2	-0.083[-0.103,-0.062]	0.010	1.362e-04	1	6225	-0.079
σ_{b0}^2	0.164[0.117,0.223]	0.026	2.781e-04	1	9710	0.164
σ_{b1}^2	0.007[0.004,0.010]	0.001	1.905e-05	1	6126	0.005

Table 2: Comparison of Bayesian model that is hierarchically centered and MLE point estimates

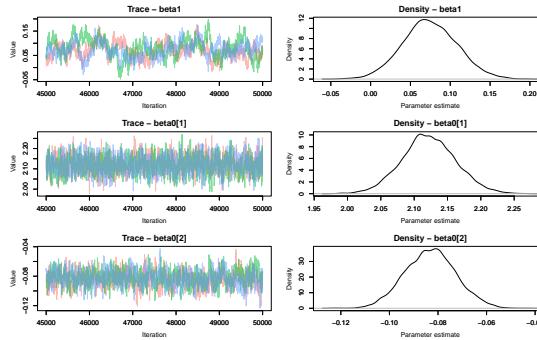


Figure 7: Trace plots for hierarchically centered model without smoothing

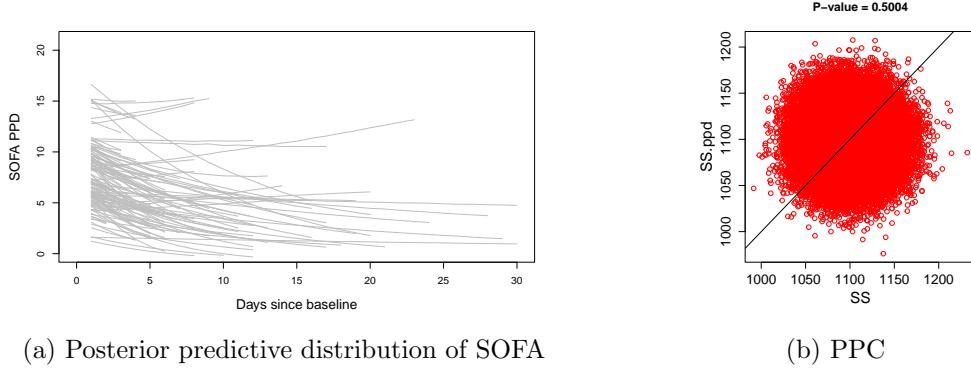


Figure 8: PPC for hierarchically centered model without smoothing

It is worth noting that a regression model considering the interaction with age and day - as the traditional formulation of longitudinal analysis does - was also run:

$$\log(sofa_{ij} + 1) = \beta_1 age_i + b_{0i} + \beta_2 age_i day_{ij} + b_{0i} + b_{1i} day_{ij}$$

However, the conditional WAIC score was 662.2; hence, the more parsimonious model was considered, accounting for the relatively low compensation in model fit that the extra parameter adds.

Based on the foregoing analysis, we conclude that a change of priors did not affect the interpretation of the model and the posterior summary results closely resemble that of the frequentist results seen earlier. β_2 is interpreted as the slope of the average line; it is the average change across all patients in SOFA score for a 1 unit change in the number of days on MV. A negative mean value for β_2 suggests that a longer stay on MV is linked with decreased SOFA score on average. The covariate *age* does not seem to be a strong indicator of SOFA score. The average SOFA score of a patient on first admittance to ICU, regardless of age, is around a value of 7. Between patients, there is substantial variance in intercepts, indicating that the starting SOFA score is likely to be patient dependent. The variance of the slopes between patients is relatively low, suggesting that most patients are likely to follow the observed downward SOFA score trajectory for increasing days on MV.

Smoothing technique As part of this analysis we also fit a mixed model-based penalized splines using a Bayesian approach. The advantages of a Bayesian approach compared to the frequentist mixed model approach include taking into account uncertainty associated with the variance components and the ability to deal with complications, such as heteroscedasticity and missing data, which cannot be handled using standard mixed model software.

With respect to mixed models, random effects are used to account for within-group dependence and splines (either penalized splines or B-splines) with mixed model

representations allow for natural extension from parametric to semi-parametric longitudinal data models. The benefit of adopting a Bayesian approach in these type of models is that they take into account the uncertainty in the smoothing parameters, which in a frequentist method would require bootstrapping.

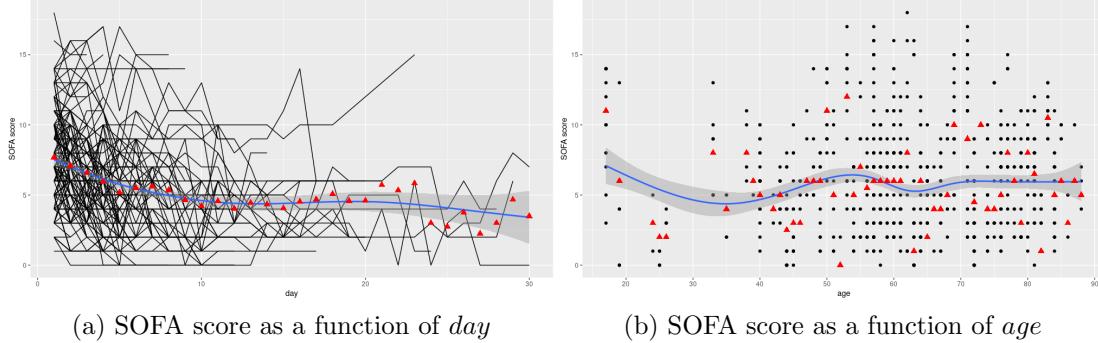


Figure 9: Average profiles of patient's SOFA score obtained by locally weighted regression using ggplot2

Using a Bayesian approach, we applied B-splines smoothing method to the individual curves of SOFA score as a function of days on MV – penalized splines heavily increased the computational cost for the MCMC simulations. The Age of the patient was not included as a covariate in this model – replaced by the resulting splines. The black lines in Figure 10 represent the ground truth values for the SOFA scores of 5 randomly selected patients, while the blue lines indicate the smoothed average curves. The number of knots used was 10 and normal distributions were assigned as priors to the random and fixed effects parameters.

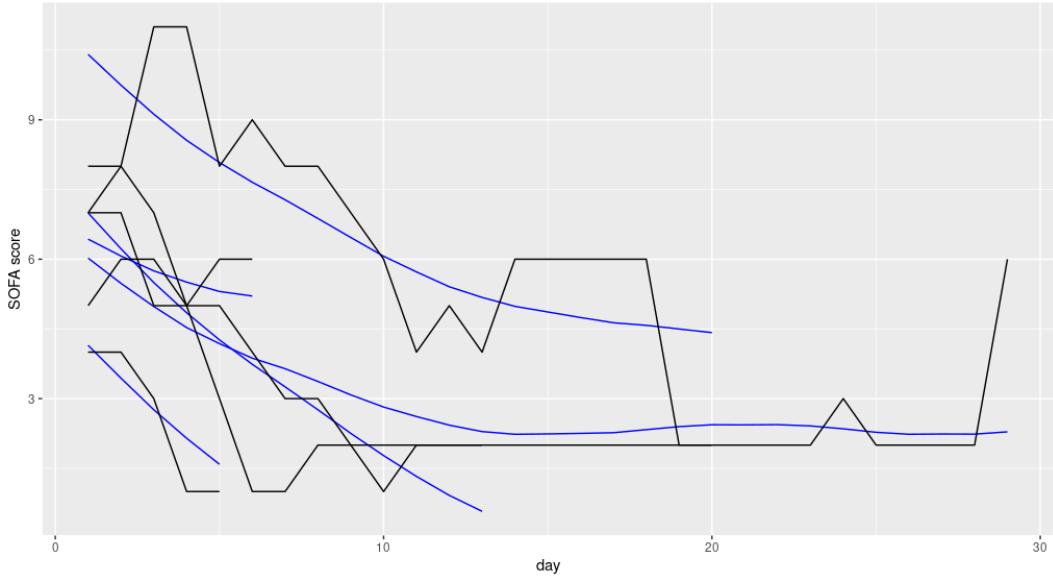


Figure 10: Fitted random slope-based penalized splines using a Bayesian approach of 5 randomly selected patients

The proposed model with B-splines can be formulated by the following equation:

$$\log(sofa_{ij} + 1) = f^*(age_i) + b_{0i} + b_{1i}day_{ij}$$

Where $f^*(.)$ is the B-spline smoothing function – considering order 3 and 10 knots. Priors for each of the 10 coefficients for the 10 knots were assumed to be independent vague normal $N(0, 10000)$ distributed; the random effects were assumed to have mean zero, since the intercept is incorporated in the splines. This model was tested using 100000 iterations, a burn-in of 50000, and thinning of 1 and 10 for convergence purposes; a thinning rate of 10 seems sufficient for convergence purposes, with most \hat{R} values equal to 1, and the highest around 1.02. Trace plots for two of the coefficients are shown in Figure 11.

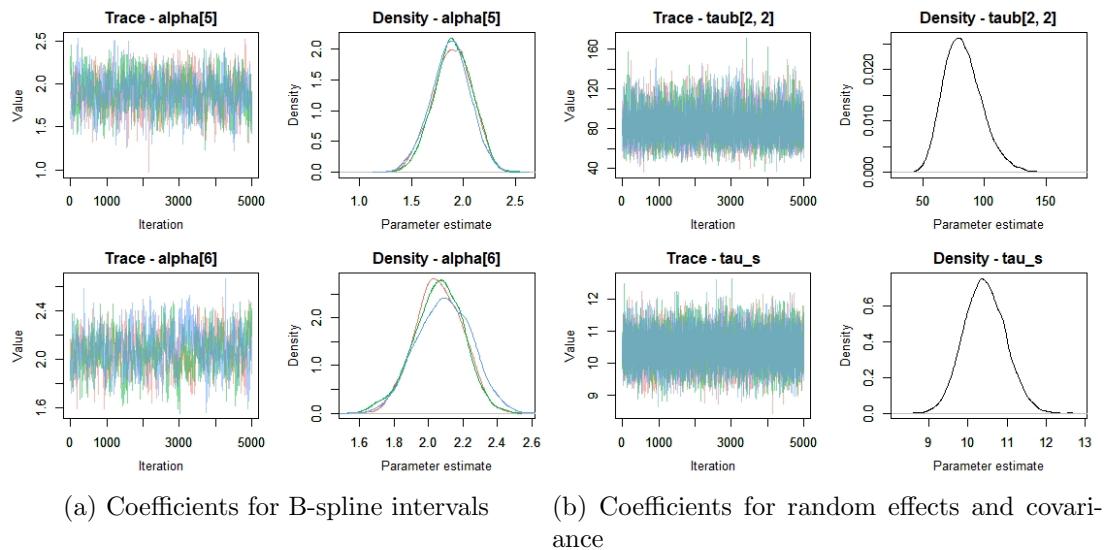
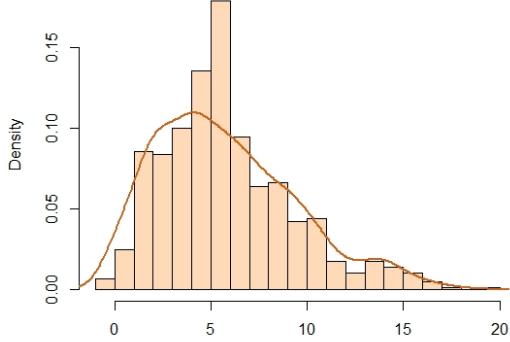


Figure 11: Traceplot of a subsample of parameters, B-spline model

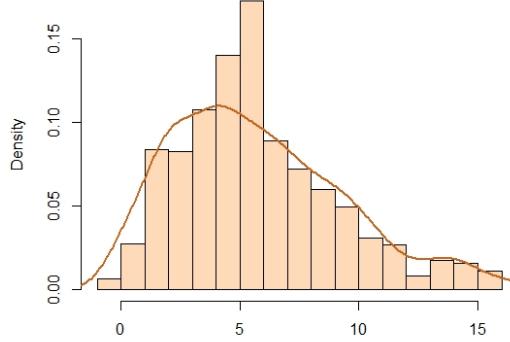
Plots for the posterior predictions for SOFA versus the original sample's density, both for the model with B-splines and for the former mixed models version, are plotted in Figure 12 – including missing value estimates for the histogram. One can observe that both models have similar performance, slightly "overpredicting" values around the mean and median of the original sample values for SOFA. The WAIC score of the outcome MCMC simulation of model is of 689.93, which is higher than the score obtained for the analogous model without splines. Hence, although smoothing techniques can be useful for capturing non-linearities, in the context of predicting SOFA scores this method is discarded in favour of a traditional mixed model approach.

B-splines based estimates vs. density of SOFA



(a) B-spline based model PPD

Longitudinal estimates vs. density of SOFA



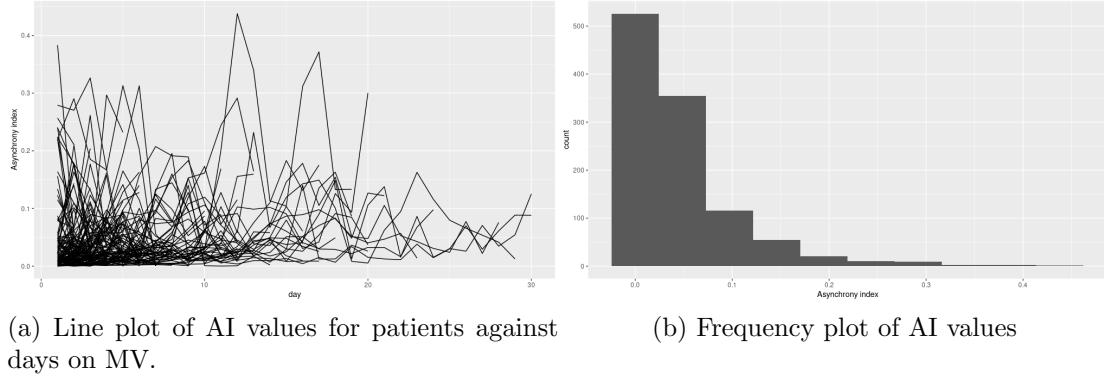
(b) PPD without B-spline smoothing

Figure 12: Histogram of posterior estimates plus density curve for original sample of SOFA, with vs. without splines

2 Analysis II

Since AI is measured as a proportion (variable di), a $Beta(\alpha(day), \beta(day))$ seems reasonable. Here $\alpha(day)$ and $\beta(day)$ are functions of day and age (possibly depending on random effects). You need to regress the mean bounded outcome (i.e., AI) as a function of the covariates of interest (similar to SOFA). Hint: Have a look at the beta regression model by Ferrari and Cribari-Neto (2004)

The Asynchrony Index (AI) is defined as the proportion of asynchronous events among the total number of ventilator cycles (range 0-1). Modelling AI is synonymous with the situation to modelling rates or proportions. Figure 13 displays the observed AI values.



(a) Line plot of AI values for patients against days on MV. (b) Frequency plot of AI values

Figure 13

To model the relationship between AI and the covariates of interest, we assume that the response follows a beta law, i.e the Beta distribution: $y \sim Beta(\alpha, \beta)$ with mean $\mu = E(y) = \frac{\alpha}{\alpha+\beta}$ and variance $Var(y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. A well known property of the beta distribution is that it is very flexible for modelling data on the standard unit interval, since the beta density can display quite different shapes, including symmetrical, skewed, uniform, roughly bell-shaped and bimodal, which depend on the values of the parameters that index the distribution. The underlying idea behind beta regression is a reparameterizing of the function of the mean of the dependent variable. The reparameterization is given by a linear predictor that is defined by regression parameters and covariates.

We use the fact that $\alpha = \mu\phi$ and $\beta = (1 - \mu)\phi$ wherein we include covariates to the mean μ and put a prior on the parameter ϕ . In particular, since μ is a bounded value between 0 and 1, we use the logit to link the mean response μ to the linear predictor; i.e. $logit(\mu_{ij}) = \beta_0 + \beta_1 age_i + \beta_2 day_{ij} + b_{0i} + b_{1i}$. After running 100,000 iterations with 50,000 burn-in, the trace plots do not look stationary yet, as displayed in Figure 14a. However, after performing hierarchical centering using the same number of iterations and burn-in, the trace plots and even the

BGR diagnostics improved a lot (potential scale reduction factor estimates are all equal to 1) (14b). Table 3 presents the posterior estimates obtained.

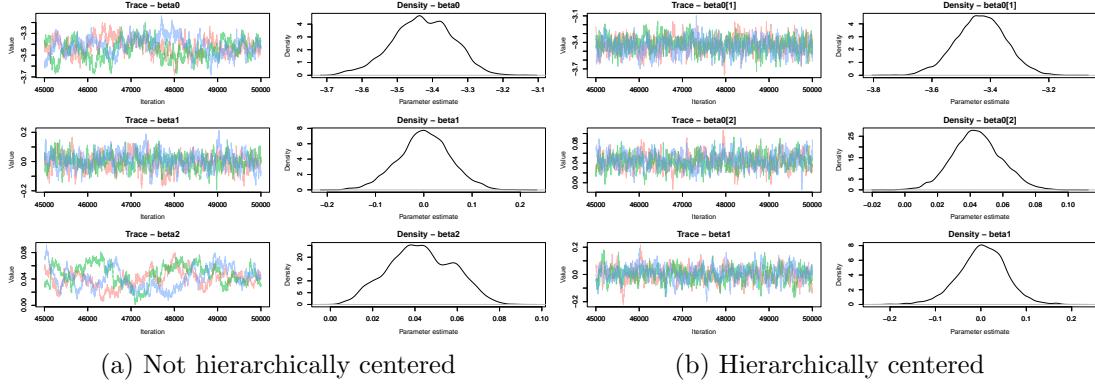


Figure 14: Trace plots for the beta regression

When the logit link function is used to transform the mean response, the regression parameters can be interpreted in terms of the odds ratio. Parameter estimation is performed by a Bayesian approach.

Parameter	mean	sd	MC error	2.5%	median	97.5%	\hat{R}	ESS
β_0	-3.430495482	0.086638002	0.0013704	-3.60129405	-3.430209870	-3.26139493	1	4256
β_1	-0.002011616	0.051549949	0.0010426	-0.10277375	-0.001778032	0.10000412	1	2626
β_2	0.043650659	0.015506627	2.747e-04	0.01315474	0.043668036	0.07380149	1	3521
$\sigma_{\beta_0}^2$	0.650519984	0.125082499	0.0014640	0.43233046	0.641333943	0.92232074	1	7787
$\sigma_{\beta_1}^2$	0.015397360	0.004232507	7.109e-05	0.00872573	0.014838563	0.02510155	1	4280
ϕ	31.398216408	1.675060490	0.019202	28.20730601	31.387256993	34.77365984	1	7615

Table 3: Hierarchically centered beta regression model

The posterior predictive distribution for AI is displayed in Figure 15. We note that there is a negative relationship between the mean response (AI value) and the general age of a patient, although this relationship is not significant, and that there is a positive relationship between the mean response and the number of days a patient spends on MV.

Density plot of B-splines based PPD

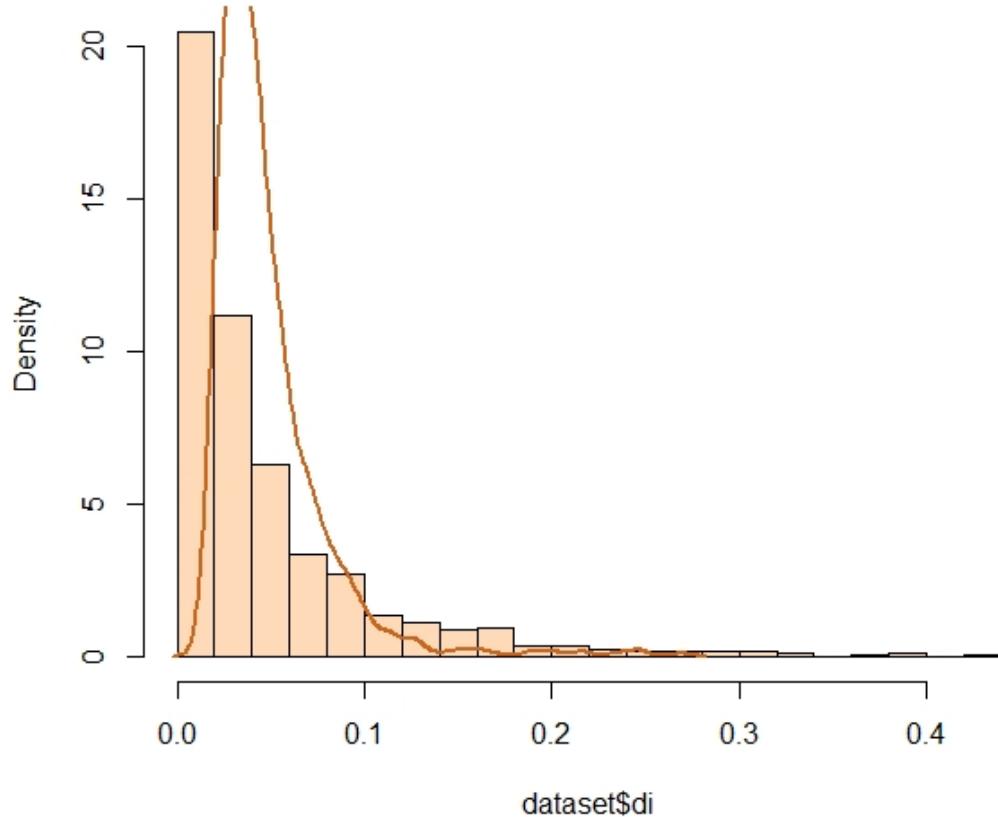


Figure 15: PPD density vs. sample histogram for AI, LMM with hierarchical centering

3 Analysis III

Simplify the response *fail* by collapsing outcomes 0 and 2 into 0. Perform a proportional hazards model with a baseline Weibull hazard function and covariates the predicted values for SOFA and AI at the global grid of observed times. Check the importance of these covariates.

Bayesian Weibull Model The proportional hazards model is often used to investigate the association between the survival time of patients and one or more covariates. In our case, we aim to model the response *time* – measured as number of days in mechanical ventilation until death in the ICU - with predicted values for SOFA and AI as the covariates of interest. The proportional hazards model will help us understand how SOFA and AI influence the rate of death in a patient at a particular point in time. This rate is regarded as the hazard rate. Replacing the baseline hazard with the Weibull hazard function gives rise to the Weibull proportional hazards model.

Let t_i be the days until death for patient $i = 1, 2, 3, \dots, 139$ and let x_i be the corresponding vector of covariates, namely predicted values of SOFA and AI. The hazard rate $h(t_i, x_i)$ for the Weibull model is a function of a baseline hazard $h_0(t_i)$ and covariates:

$$h(t_i; x_i) = h_0(t_i) \times h_i(t_i) = h_0(t_i) \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (3)$$

where $h_0(t_i) = \rho t_i^{\rho-1}$ and $\boldsymbol{\beta}$ contains an intercept term. Therefore, we have that

$$h(t_i, x_i) = \rho t_i^{\rho-1} \exp(\mathbf{x}_i^T \boldsymbol{\beta})$$

$$t_i \sim \text{Weibull}(\rho, \lambda_i) \quad (4)$$

$$\rho \sim \text{Gamma}(0.0001, 0.0001)$$

$$\beta_i \sim N(0, 10000), \quad i \in \{0, 1, 2\}$$

where $\lambda_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ and $h_0(t_i)$ is regarded as a baseline hazard, while $h_i(t_i)$ is a non-baseline component which will be the focus of our modelling since it can include the covariates SOFA and AI. Given that the response is right censored (1: dead in the ICU; 0: censored at day 30), we make use of Nimbles' censoring function. The data vector for \mathbf{t} is given *NaN* values (indicating missing data) for any censored $\mathbf{t}[i]$ entries. In Nimble, the vector `censored[i]` should be given as data with a value of 1 if $\mathbf{t}[i]$ is right-censored and 0 if it is observed. Therefore we swap the values such that 0: dead in the ICU and 1: censored at day 30. Furthermore, we set the vector `c` to give the censoring times corresponding to censored entries and a value above the observed times for uncensored entries (e.g Inf).

Beginning with the formulation $\lambda_i = \exp(\beta_0 + \beta_1 SOFA_i + \beta_2 AI_i)$ for the parameter λ_i , we standardize the measures of SOFA and AI, then run 100,000 iterations with 50,000 burn-in and 3 chains; the resulting trace, running mean, and cross-correlation plots are displayed in Figure 16. We observe that the chains for β_1 and β_2 stabilize much quicker than the ones for β_0 and ρ . This is explained mainly by the high cross-correlation between parameters β_0 and ρ — as observed in Figure 16c.

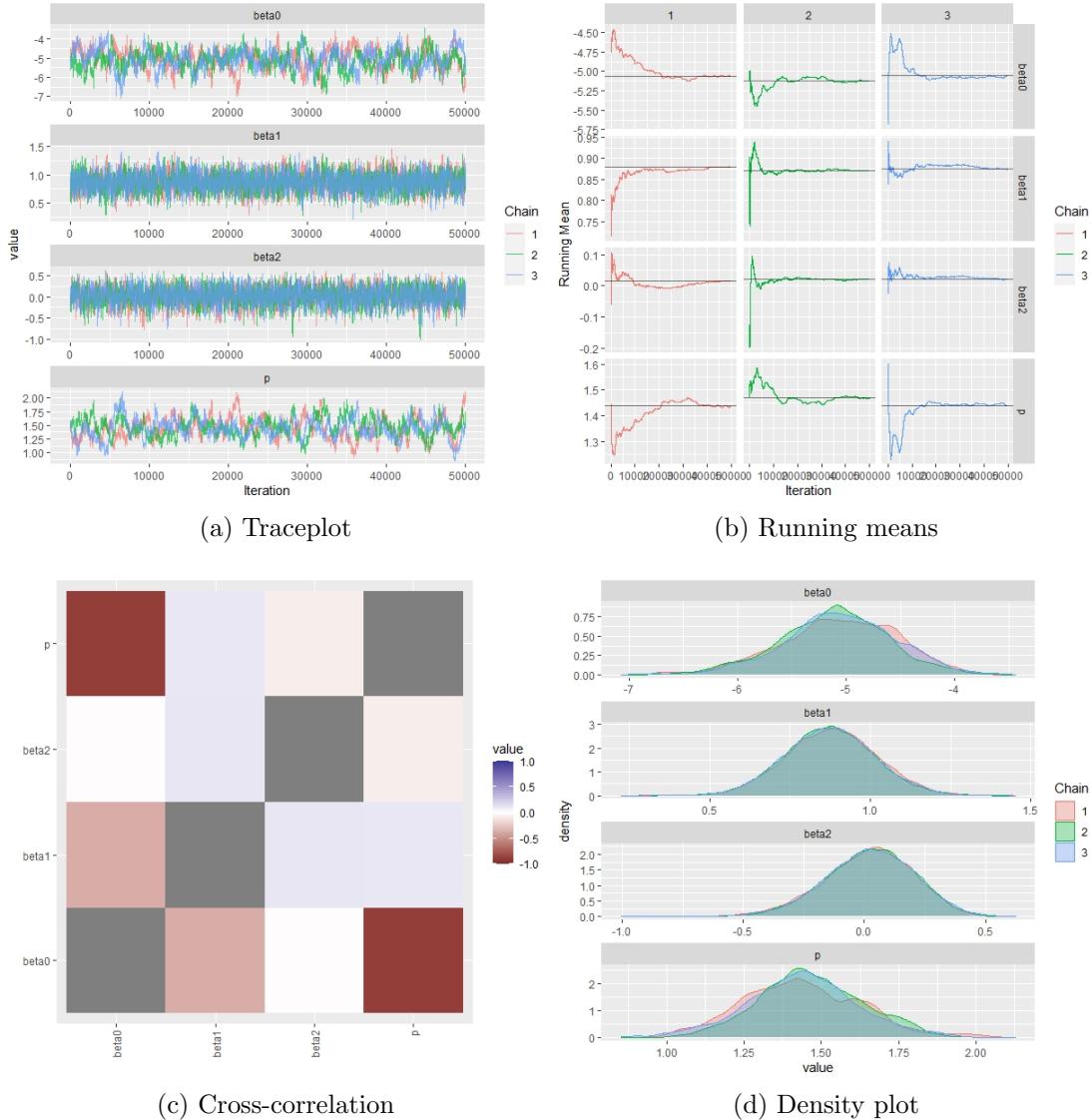


Figure 16: Trace, running means, cross-correlation and density plots for Weibull model, 100,000 iterations with 50,000 burn-in, no thinning

Nonetheless, although convergence is not perfect, it seems sufficient, as observed in the resulting \hat{R} for all four parameters β_0 , β_1 , β_2 and ρ being between 1.00 and 1.01, and as observed in the MC errors all being below 0.05 — as displayed later. Both thinning the chain (with 5 and 10) and increasing the burn-in to 80,000

worsened the convergence of parameters; changing the prior distribution for the shape parameter to a uniform distribution $\rho \sim U(0, 1000)$ slightly improved \hat{R} scores for the four parameters, yet did not significantly alter the resulting coefficients' means or standard deviations. Hence, we use the former results for further analysis.

Resulting coefficients are displayed in Table 4, along with MLE estimates for comparison (Note: from the *survreg* loglinear form of the proportional hazards model, we converted the coefficients into their corresponding hazard coefficients which are shown in the table). It is observed that a Bayesian analysis results in the (standardized) SOFA score — as represented by β_1 — having a significantly positive contribution to the survival time of patient. This means that an increase in the severity of illness (as measured by SOFA) translates to an increased hazard or risk for patients in the ICU with an MV.

Parameter	Pos mean[2.5%,97.5%]	sd	MC error	Rhat	ESS	MLE
β_0	-5.236 [-6.416, -4.154]	0.571	0.037	1.01	238	-5.08056
β_1	0.874 [0.595, 1.155]	0.142	0.003	1.00	2056	0.8701031
β_2	-0.202 [-0.647, 0.168]	0.209	0.005	1.00	1870	0.04388807
ρ	1.539 [1.128, 1.986]	0.214	0.019	1.01	121	1.472754

Table 4: Comparison of Bayesian model for survival time vs. MLE point estimates, Weibull distribution

Under the assumption of proportional hazards, we can interpret the hazard ratios of covariates as multiplicative effects on the hazard. Holding AI constant, SOFA has a hazard ratio of $\exp(\beta_1) = 2.401$, indicating a strong relationship between SOFA score and an increased risk of death, i.e. a one-unit increase in SOFA score implies 2.401 times more risk for a patient in MV. In other words, a higher SOFA score is associated with poor survival. On the other hand, the covariate AI does not make a significant contribution to the difference in hazard ratio. This is evident by the obtained credible interval for the parameter which includes 0.

4 Conclusion

Considering that the aim of the research is to assess the impact of the Asynchrony Index (AI) and Sequential Organ Failure Assessment (SOFA) on a patient's survival, we performed two longitudinal analyses for each covariate's respective distribution and fit, then used those joint results to fit a proportional hazards model for patients' survival time.

Regarding SOFA, covariates *age* and *day* (under ICU treatment) were considered for the longitudinal analysis. Considering the skewness of SOFA's distribution, a log-transformation of it — accounting for zero values by adding 1 to each score - plus a normal distribution of this transformation was deemed sufficient; missing values were assumed a prior near the log-mean, and predicted within each iteration of the MCMC chain.

Random effects both for the intercept and for the slope — the latter defined by time under ICU -, with Wishart priors for the covariance matrix, considerably improved the performance of the model and its overall fit. Complementarily, smoothing was attempted for the covariate *age* via B-splines in order to account for non-linearities; however, inclusion of splines did not noticeably improve the fit of the model, thus was discarded. Results show that a longer stay in ICU is linked with a decreased SOFA score on average, thus lower severity of illness; age does not seem to significantly relate to SOFA scores.

As for the modeling of AI, a beta distribution was fitted — since this is measured as a proportion among the total number of ventilator cycles. To account for individual differences in response, random effects for the intercept and for the slope — the latter also defined by the time variable - were included. Results suggest a positive relationship between the amount of asynchronous events under mechanical ventilators (MV) and the number of days spent under MV; age, although displaying a positive relationship, is not significant.

Finally, regarding the distribution for survival time of patients, a proportional hazards model with baseline Weibull hazard was implemented, using as covariates the fitted scores for SOFA and AI for the last day under ICU. We observe that a higher SOFA score, thus the severity of illness, is associated with higher risk for the patients, while AI does not by itself have significant impact on the hazard.

References

Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3): 515 – 534, 2006. doi: 10.1214/06-BA117A. URL <https://doi.org/10.1214/06-BA117A>.