Right censured repeated measures over time
How to analyze change in pain thresholds

Joseph Mietkiewicz

Göteborg University
2021

Contents

I would like to thank my advisor Anna Ekman
in particular for here supervision and
guidance throughout all this master thesis
I would also like to thank Petter Mostad
for his very helpful comments on the project

1 Introduction

This project studies a model for data from pain threshold research. It is based on the study "Pain intensity and pressure pain thresholds after a light dynamic physical load in patients with chronic neck-shoulder pain" by A. Grimby-Ekman, C. Ahlstrand, B. Gerdle, B. Larsson and H. Sandén (2020) [1]. The idea of this research is to measure the pain of sick patients over time after physical exercises . Pressure pain threshold was used to measure pain. Pressure with an instrument is performed on the patient and when they feel pain, the pressure is reported. The pressure is applied to six different locations: Trapzeius 1, 2 and 3 on the right and left side. It is measured at four different times, before a physical exercise and then at 3 times afterwards. To avoid bruising and pain, the pressure pain thresholds are capped at a maximum value (700 kPa). If a patient did not feel pain after this maximum, the 700 value is reported and treated as the true measurement. Since the measurement is repeated over time and a natural random effect occurs between the patient, a linear mixed model is used. We will examine this linear mixed model in detail. The parameter estimation of the study does not take into account the censoring of the data. The censored data is considered as true data point. The idea of this project is to use the Tobit model for censored data to improve the parameter estimation. The first idea is to maximize a likelihood function that takes censoring into account with a standard optimization algorithm. This algorithm is studied in detail in another project. Instead, a Bayesian approach is studied. The Gibbs sampling is a natural candidate to deal with censored data. The idea of this algorithm is to redraw the censored data with a truncated normal distribution at each iteration. The Expectation Maximization (EM) have an other approach. The algorithm average according to the truncated normal distribution the censored measurement at each iteration. After the study of these algorithms, a comparison between taking into account and not taking into account the censorship is made. To study the impact of censoring, we ran the algorithms on simulated data with different levels of censoring. The finale goal is to provide a practical answer for handling data from pain threshold research.

2 The pain Threshold research

2.1 The study

Chronic musculoskeletal pain is a common clinical condition that brings patients to the doctor's office and is a major cause of disability and reduced work capacity. Clinical experience suggests that some patients with chronic musculoskeletal pain may experience increased pain intensity the day after even mild physical exertion. It is important to take this into account in the evaluation of work capacity. The idea is to study the pain felt over time after a physical effort. Two groups are studied: the patient with chronic musculoskeletal pain and a control group. The study measures the pain pressure threshold to evaluate the pain. The measurement was done in 4 different times. Before the exercise, 15 min, 1 hour and 1 day after. And on 6 different locations: trapezius 1,2,3 left and right.

The goal of the study was to investigate the development of pain intensity and pressure pain thresholds during and 24 h after a light dynamic physical load among patients with chronic neck-shoulder pain.

The result of the study is that the patient has a pain threshold that decreases right after the physical exercise to return to the same level after one hour and after one day. The control group shows an increase in pain threshold for the first three measurements and stay at the same level for the last measurement.

2.2 The data

Pressure pain threshold (PPT) was measured by a hand held electronic pressure algometer and measured in a standardized manner. The contact area was 10 mm and the pressure was applied at a rate of 30 kPa/second. The participants were instructed to mark the PPT by pressing a signal button when they felt the first sensation of pain. At a maximum value of 700 kPa the measurement was interrupted to avoid bruising and soreness induced by the measurement method. The people involved in the experiment are summarized in the table below.

	Men	Women	total
Patient	5	21	26
Control	6	7	12
total	11	28	39

The following graph represents the pain thresholds for the sick group (in black) and for the control group (in red). The first 6 graph is for women and the next one for men. Each graph represents the pain threshold reported at a given time. The three top graphs correspond to the three trapezius locations on the left and the three bottom graphs on the right. It is clear that each person has a different pain threshold. A random effect is needed to model the data. It appears that a different pattern emerges for the left and right measurements. The bolded lines represent the mean. On average, the control group has a higher pain threshold, and is therefore more resistant to pain. The impact of censoring is clearly visible. Some pain thresholds at 700 appear for all 4 measurements. In the data set 4.8% of the data are censored.

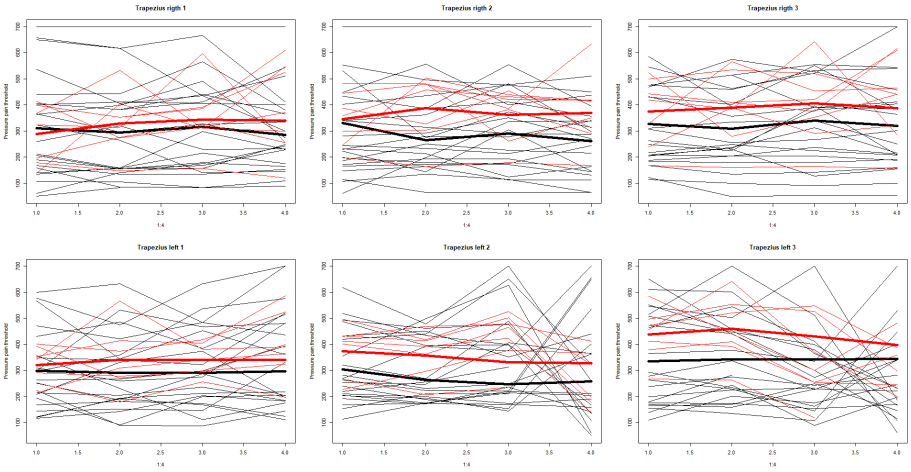

Figure 1: Individual lines and mean threshold at each time point for women for the two group. The top three graph are the 3 right localisation and the 3 down the left. In red the control group

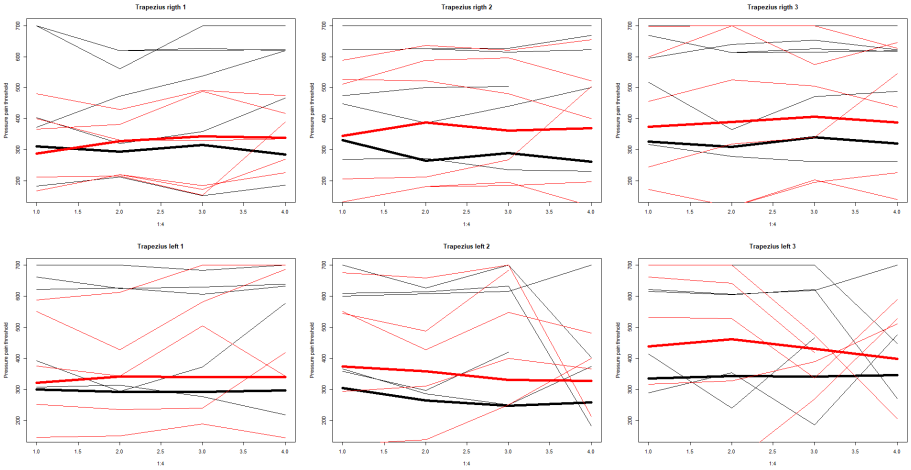

Figure 2: Pain thresholds for men for the two group

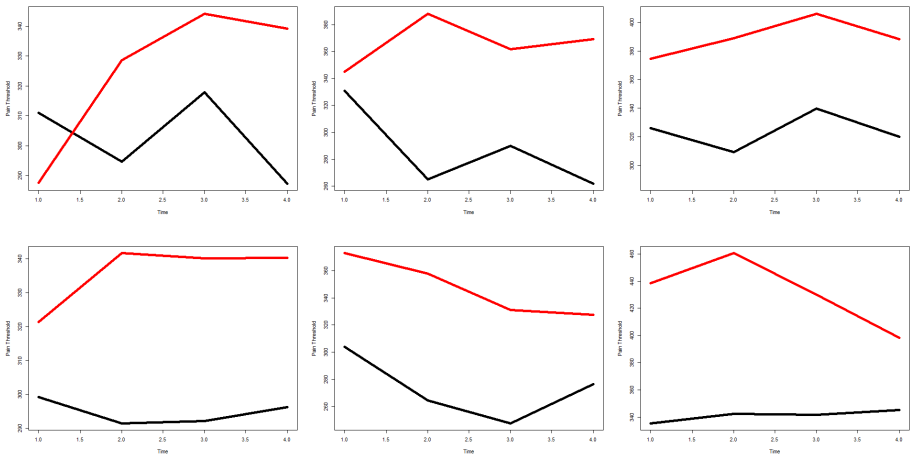

Figure 3: Average pain thresholds of women for both groups

The average of the main pain threshold already shows a pattern. The decrease after exercise is already present on the right side. The left side shows a different pattern. The idea is build a model to have a more precise answered to the change of pain threshold over time.

3 The model for pain thresholds research

Because the measurements are taken over time and on different individuals, the data are modeled with a linear mixed model. To model the random effect between patients, a random intercept is used. The left and right measurements also lead to another random effect. We first describe the mathematical assumption of the model and then the parameters used in this specific study.

3.1 Linear mixed model for longitudinal data

The linear mixed model is a regression model. It allowed to estimated both a fixed and random effect.

The linear mixed model use in the project satisfies:

$$\begin{cases} Y = X\beta + W\mu + \epsilon \\ \text{or} \\ Y_i = W_i'\beta + W_i\mu_i + \epsilon_i \text{ for i=1,...,N} \\ \text{or} \\ y_{its} = x_{its}'\beta + \gamma_1 + \gamma_s + \epsilon_{its} \text{ for i=1,...,N, t=1,...,T and s=1,2} \end{cases}$$

The Y vector is the reponses variable (the observation) of length N×T. The X matrix is the designe matrix of the fixed effect. W the designe matrix of the random effect. μ the vector of the random effect and ϵ the random error vector. W

$\mu_i = (\gamma_1, \gamma_2, \gamma_3)'$ with $\gamma_s \sim N(0, \sigma_{\gamma_s})$ and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. Let μ_i ∼N(0,R) and

We assume strict exogeneity:

$$\begin{aligned} E[\epsilon_{it}|X] &= E[\mu_i|X] = 0 \\ E[\epsilon_{it}^2|X] &= \sigma_\epsilon^2 \\ E[\gamma_j^2|X] &= \sigma_{\gamma_j}^2 \\ E[\epsilon_{it}\gamma_j|X] &= 0 \text{ for all i, t and j} \\ E[\epsilon_{it}\epsilon_{js}] &= 0 \text{ if t} \neq s \text{ or } i \neq j, \\ E[\gamma_i\gamma_j|X] &= 0 \text{ if } i \neq j \end{aligned}$$

$$\eta_{its} = \epsilon_{it} + \gamma_1 + \gamma_s$$

$$\begin{aligned} E[\eta_{its}^2|X] &= \sigma_\epsilon^2 + \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 \\ E[\eta_{its}\eta_{ins}] &= \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 \ , \ n \neq t \\ E[\eta_{its}\eta_{inv}] &= \sigma_{\gamma_1}^2 \ , \ n \neq t \ , \ s \neq v \\ E[\eta_{its}\eta_{jnv}] &= 0 \text{ for all t,n,s and v if } i \neq j. \end{aligned}$$

For the T observations for the person i, $\Sigma_i = E[\eta_i \eta_i' | X]$. Then

$$\Sigma_i = \begin{pmatrix} \sigma_\epsilon^2 + \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 & \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 & \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 & \cdots \\ \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 & \sigma_\epsilon^2 + \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 & \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 & \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 & & \sigma_\epsilon^2 + \sigma_{\gamma_1}^2 + \sigma_{\gamma_s}^2 \end{pmatrix}.$$

Or

$$\Sigma_i = W_i R W_i' + D$$

The disturbance covaraince matrix for all the observation is:

$$\Omega = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \Sigma \end{pmatrix}.$$

These assumptions mean that the measures between individuals are independent and have the same variance structure.

3.2 The model of the study

We will now specify the parameters used in the study. The model is used in the study is:

$$Y = X\beta + W\mu + \epsilon$$

The parameters are time, location, group, time× group and gender.All variables are categorical. N is the number of people in the study and T is the number of measurements for each person. In this study N=39 and T=24 (3 measurement on each side for the 4 different time)

Y is the vector of outcomes of the $N \times T$ measures. Beta is the fixed effect vector of length 11 (The 4 times are coded with 3 dummy variables, the location by 2, the group by 1, the interaction between time and group by 3 and the gender by 1 dummy variable). The random effect is a vector of length 3×N (to model the random intercept and the two random effects of the left and right measurement) and the random error is a vector of length $N \times T$.

The matrix X is a ($N \times T$, 11) matrix :

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrix W is a $(N \times T, 3N)$ matrix:

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}; \quad B = \begin{pmatrix} A \\ A \\ A \\ A \end{pmatrix}; \quad W = \begin{pmatrix} B & & (0) \\ & \ddots & \\ (0) & & B \end{pmatrix}$$

4 Linear mixed tobit model

The linear mixed model fits the way the data are collected but does not account for censoring. We first show the problem that can arise if all the data are treated as real measurements, and then we study the Tobit model to handle censoring.

4.1 The issue of not taking into account censoring

The most simple way to find the ML is the least square estimate. The formula for the least square estimate is similar to the linear model:

$$\hat{\beta} = (X'X)^{-1}X'(Y - W\mu)$$

Let Y^* the true unknown data set without the censoring.

$$E[\hat{\beta}] = E[(X'X)^{-1}X'(Y - W\mu)] = (X'X)^{-1}X'(E[Y]) \leq (X'X)^{-1}X'(E[Y^*] = (X'X)^{-1}X'X\beta = \beta$$

So the β estimate is biased and on average underestimate.This can lead to a poor regression.

An other issue cause by censoring is with the assumption of the model. The normality assumption is no respected with the censored points.The two following graph are the QQ plot of the conditional residual for Gibbs sampling with censored data considered as true value. The conditional residual is defines as $R = Y - X\beta - W\mu$.

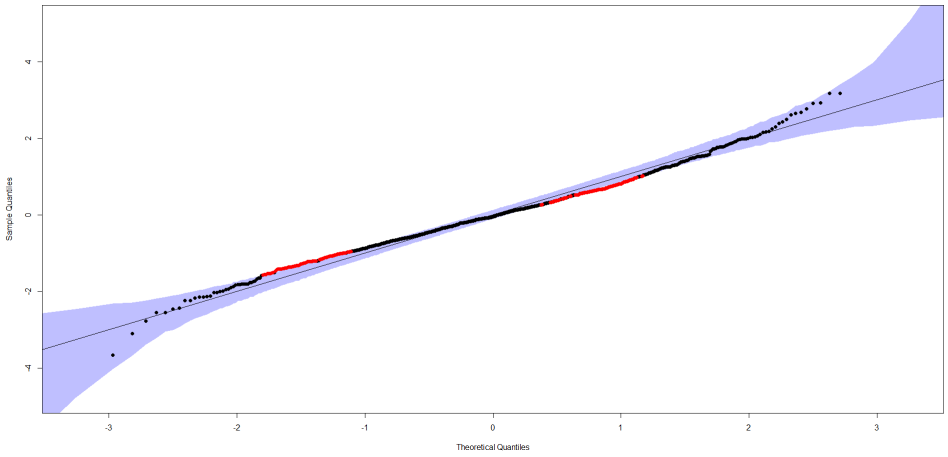

Figure 4: QQ plot without censoring taken into account

The two algorithm give similar result. A number of point are out of the 95% tolerance band. This question the normality assumption of the residual.

The fitted value versus the conditional residual is use to detect outlying observation. The censorship can be detected in this plot. The line is 700-fitted value.

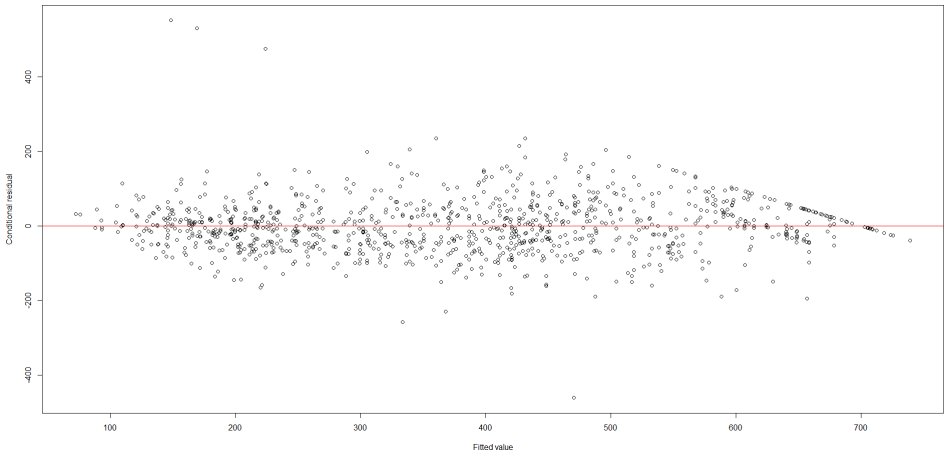

Figure 5: Fitted value versus Conditional residual

4.2 Longitudinal Tobit Model

The Tobit model is a regression model where the dependent variable is subject to a certain constraint. The Tobit model was introduced by Tobin (1958)[2]. It was originally developed to model household expenditures. He developed a regression model that takes into account the fact that expenditures cannot be less than zero. This type of data can be called left-censored data. Amemyia (1985) classified the Tobit model into 5 types, each differing in the form of the likelihood. In this project, we are interested only in the type 1 Tobit model with right censoring data. This model is described as follows:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* < l \\ l & \text{else} \end{cases}$$

With y_i the response variable and y_i^* the true and unknown value (if the censoring had not occurred)

The longitudinal Tobit model implement the censoring in the classical linear mixed model. The model is described in [3]
The mixed Tobit model is defined here as:

$$y_{its}^* = x_{it}'\beta + \gamma_1 + \gamma_s + \epsilon_{its}.$$

$$y_{its} = \begin{cases} y_{its}^* & \text{if } y_{it}^* \leq l \\ l & \text{else.} \end{cases}$$

With $\gamma_j \sim N(0, {\sigma_\gamma}_j^2)$ and $\epsilon_{its} \sim N(0, \sigma_\epsilon^2)$. So $y_{its}^* \sim N(x_{its}'\beta, {\sigma_\gamma}_1^2 + {\sigma_\gamma}_s^2 + \sigma_\epsilon^2)$

The density function is:

$$f(y_{ijs} = l) = P(y_{ijs}^* < l) = F(y_{ijs}^* = l)$$
$$f(y_{ijs}) = f(y_{ijs}^*) \quad \text{for} \quad y_{ijs} < l.$$

The likelihood of i measurement is:

$$L_i = \int_{b_i} \prod_j f(y_{ij}) N(b_i; 0, D) db_i.$$

The likelihood for all cases is:
$L = \prod_i L_i$

So for T_i observation belonging to the i individual we have the likelihood contribution :

$$L_i = \int_{-\infty}^{\infty} \Big(\prod_{t=1}^{T_i} [\frac{1}{\sigma_\epsilon} \phi(\frac{y_{it} - \beta' x_{it} - \mu_i}{\sigma_\epsilon}]^{d_{it}} \big[\Phi(\frac{\beta' x_{it} + \mu_i - l}{\sigma_\epsilon})\big]^{1-d_i t} \Big) \phi(\frac{\mu_i}{\sigma_\mu}) d\mu_i.$$

$d_{it} = 1$ for uncensored observation and 0 else. The log likelihood for the whole sample is:

$$L = \sum_{i=1}^{N} \log(L_i).$$

This likelihood function can be maximized with a standard optimization algorithm using Gauss-Hermit quadrature to compute the integral, but this method is not studied here. Instead, we use EM and Gibbs sampling. Both of these algorithms use the truncated normal distribution to handle censoring. A left-truncated normal distribution has the following density for x$\geq$ l, l being the limit.

$$f(x, m, \sigma^2) = \frac{N(x, m, \sigma^2)}{1 - \Phi(\frac{l-m}{\sigma})}$$

With Φ the standardized normal distribution function. The following graph represents the truncated normal distribution with the limit set at 0.3 in blue and the normal distribution with mean 0 and variance 1. The Gibbs sampling draws the censored data belonging to this distribution with the right parameters. The EM algorithm integrates on the censored variable with this distribution.

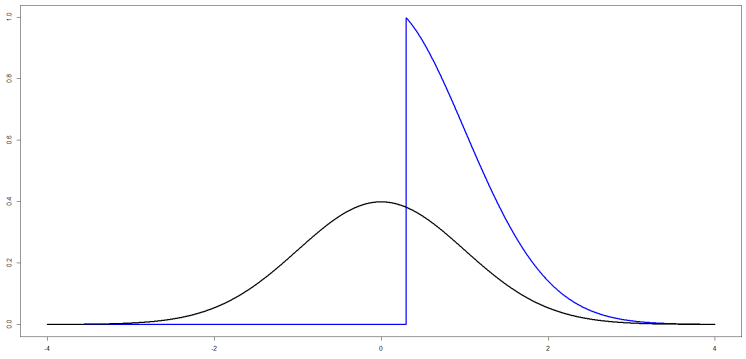

Figure 6: Normal and left truncated normal distribution

5 The Bayesian approach

5.1 Gibbs sampling

The Gibbs sampler is well suit for missing data. The censored data can simply be redraw according to the truncated normal distribution at each iteration. Lee, Seung-Chun and Choi, Byongsu (2014) [4], Bruno (2004) [3]

for i=1,...,N , t=1,...,T and s=1,2:

$$y^*_{its} = x'_{its}\beta + \gamma_1 + \gamma_2 + \epsilon_{its}.$$

$$y_{its} = \begin{cases} y^*_{its} & \text{if } y^*_{its} \leq l \\ l & \text{sinon.} \end{cases}$$

With matrix notation

$$y^* = X\beta + W\mu + \epsilon.$$

The posterior distribution is:

$$p(\beta, \mu, \sigma^2_\mu, \sigma^2_\epsilon | y^*) \sim p(y^*|\beta, \mu, \sigma^2_\epsilon)p(\beta)p(\mu|\sigma^2_\mu)p(\sigma_\mu{}_{\gamma 1}{}^2)p(\sigma_\mu{}_{\gamma 2}{}^2)p(\sigma_\mu{}_{\gamma 3}{}^2)p(\sigma^2_\epsilon).$$

This mixed Tobit model give some non-linear panel data. To solve this issue we use data augmentation strategies: y^*_{it} has truncated normal distribution on [l,∞(. Otherwise, it has a degenerating distribution at y_{it}. It enable us to recursively simulate the entire posterior distribution of the parameters. The pdf of the truncated normal distribution is

$$y^*_{its} \sim \frac{N(x'_{its}\beta + \gamma_i + \gamma_s, \sigma^2_\epsilon)}{1 - \Phi(\frac{l - x'_{its}\beta}{\sigma_\epsilon})}$$

The prior distribution of β and σ_ϵ in absence of prior information is $p(\beta, \sigma^2_\epsilon) \sim \frac{1}{\sigma^2_\epsilon}$ and $\mu_i \sim N(0, R_T)$ with $R_T = diag(\gamma_1, \gamma_2, \gamma_3)$ and $\mu \sim N(0, R)$ with a non informative prior for $\sigma^2_{\gamma_j}$, $p(\sigma^2_{\gamma_j}) \sim \frac{1}{\sigma_{\gamma_j}}$ for j=1,...,3.

We sample Y with the previous formula then we estimate β, σ_ϵ and σ_μ with the sample Y.

Then the algorithm is:

$y^*_{its}|, \mu_i, \sigma_\mu, \sigma_\epsilon \sim \frac{N(x'_{it}\beta+\gamma_i+\gamma_s, \sigma^2_\epsilon+{\sigma_\gamma}^2_1+{\sigma_\gamma}^2_s)}{\Phi(\frac{l-x'_{it}\beta}{\sqrt{\sigma^2_\epsilon+{\sigma_\gamma}^2_1+{\sigma_\gamma}^2_s}})}$

$\beta|\epsilon, \sigma^2_\epsilon, y^* \sim N((X'X)^{-1}X'(y^* - W\mu), \sigma^2_\epsilon(X'X)^{-1})$

$\mu|\beta, \sigma^2_\mu, \sigma^2_\epsilon, y^* \sim N((W'W + \sigma^2_\epsilon/R)^{-1}W'(y^* - X\beta), \sigma^2_\epsilon(W'W + \sigma_\epsilon/R)^{-1}$

$\sigma^2_{\mu_{\gamma 1}}|\mu \sim IG((N-1)/2, \mu'_{\gamma 1}\mu_{\gamma 1}/2).$

$\sigma^2_{\mu_{\gamma 2}}|\mu \sim IG((N-1)/2, \mu'_{\gamma 2}\mu_{\gamma 2}/2).$

$\sigma^2_{\mu_{\gamma 3}}|\mu \sim IG((N-1)/2, \mu'_{\gamma 3}\mu_{\gamma 3}/2).$

$\sigma^2_\epsilon|\beta, \mu, y^* \sim IG(NT/2, 1/2(y^* - X\beta - W\mu)'(y^* - X\beta - W\mu))$

5.2 Henderson's mixed model

The Gibbs sampler described previously can be computationally heavy. To reduce the complexity of the algorithm the Henderson's mixed model can be used. This improvement is described in Cs Wang, Jj Rutledge, D Gianola (1994)[5].
As previously:

$$\epsilon \sim \mathrm{N}(0, I\sigma^2_\epsilon)$$

$$\mu \sim \mathrm{N}(0, R)$$

The Henderson's mixed model is described by :

$$Z\hat{\theta} = b$$

With: $Z = \frac{1}{\sigma^2_\epsilon}\begin{pmatrix} X'X & X'W \\ W'X & W'W + R^{-1}\sigma^2_\epsilon \end{pmatrix}, \hat{\theta} = \begin{pmatrix} \hat{\beta} \\ \hat{\mu} \end{pmatrix}, b = \sigma_\epsilon^{-2}\begin{pmatrix} X'Y \\ W'Y \end{pmatrix}.$

The solution to the Henderson's "mixed model equations" $\hat{\beta}$ and $\hat{\mu}$ are the best unbiased estimates and predictor for β and μ respectively

let

$$\theta' = (\beta', \mu_1..., \mu_{3*N}) = (\theta_1, ..., \theta_M).$$

With M=11+3×N (μ_i is a 3 dimensional vector and β a 11).
and

$$\theta'_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ...\theta_M)$$

Z=$\{z_{ij}\}$ for i,j=1,...,M and b=$\{b_i\}$, i=1,...,M.

As prove in [5], the conditional posterior distribution of each of the θ_i is a normal with mean and variance $\tilde{\theta_i}$ and $\tilde{v_i}$:

$$\theta_i|Y, \theta_{-i}, \sigma_\epsilon \sim \mathrm{N}(\tilde{\theta_i}, \tilde{v_i}).$$

With $\tilde{\theta_i} = (b_i - \sum_{j=1, j\neq i}^{M} z_{ij}\theta_j)/z_{ii}$ and $\tilde{v_i} = \sigma_\epsilon^2/w_{ii}$

This method is more computationally efficient since we do not invert any matrix and only compute scalars.

μ_i is a three dimensional vector $\mu_i = (\gamma_1, \gamma_2, \gamma_3)$. With the notation $\mu_{\gamma 1} = (\gamma_{11}, ..., \gamma_{N1})$, $\mu_{\gamma 2} = (\gamma_{12}, ..., \gamma_{N2})$ and $\mu_{\gamma 3} = (\gamma_{13}, ..., \gamma_{N3})$. And $\gamma_{ij} \sim \mathrm{N}(0, \sigma^2_{\mu_{\gamma j}})$

The new Gibbs sampling is :

$y^*_{its}|, \mu_i, \sigma_\mu, \sigma_\epsilon \sim \frac{N(x'_{it}\beta + \gamma_i + \gamma_s, \sigma^2_\epsilon + {\sigma_\gamma}_1^2 + {\sigma_\gamma}_s^2)}{\Phi(\frac{l - x'_{it}\beta}{\sqrt{\sigma^2_\epsilon + {\sigma_\gamma}_1^2 + {\sigma_\gamma}_s^2}})}$.

$\theta_i|Y, \theta_{-i}, \sigma_\epsilon \sim \mathrm{N}(b_i - \sum_{j=1, j\neq i}^{M} Z_{ij}\theta_j)/z_{ii}$, $\sigma^2_\epsilon/z_{ii})$.

$\sigma^2_{\mu_{\gamma 1}}|\mu \sim IG((N-1)/2, \mu'_{\gamma 1}\mu_{\gamma 1}/2)$.

$\sigma^2_{\mu_{\gamma 2}}|\mu \sim IG((N-1)/2, \mu'_{\gamma 2}\mu_{\gamma 2}/2)$.

$\sigma^2_{\mu_{\gamma 3}}|\mu \sim IG((N-1)/2, \mu'_{\gamma 3}\mu_{\gamma 3}/2)$.

$\sigma^2_\epsilon|\beta, \mu, y^* \sim IG(NT/2, 1/2(y^* - X\beta - W\mu)'(y^* - X\beta - W\mu))$.

6 Maximum likelihood estimate

After a Bayesian approach to estimate the model parameters, we study a maximum likelihood algorithm. The specificity of this algorithm is that it uses the maximization of the linear mixed model without censoring. Censoring is only taken into account at each iteration of the algorithm. So that the censored data are eliminated from the calculation, we integrate on them . We first describe the likelihood function and the algorithm without censoring and modify it in a third part.

6.1 Likelihood function

We find here the formula of the parameters to maximize the likelihood for the linear mixed mo

$$Y = X\beta + W\mu + \epsilon$$

Y$\sim$N($X\beta$,WRW'+D) and $Y|\mu \sim N(X\beta + W\mu, D)$

If μ is known, the log likelihood function is:

$$\begin{aligned} LL(\theta) &= \log(p(y, \mu; \theta) \\ &= \log(p(y|\mu; \theta) + \log(p(\mu; \theta)) \\ &= \log(p(y|\mu; \beta, \sigma_\epsilon) + \log(p(\mu; \sigma_\mu)). \end{aligned}$$

Then σ_ϵ, β minimizes:

$$-\log(p(y|\mu, \theta)) = n\log(\pi\sigma^2) + \frac{||Y - X\beta - W\mu||^2}{\sigma_\epsilon}.$$

and R minimizes:

$$-2\log(p(\mu; \sigma_\mu)) = N\log(2\pi) + N\log(R) + \sum_{i=1}^{N} \mu_i R^{-1} \mu_i''.$$

So:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'(Y - W\mu) \\ \hat{R} &= \frac{1}{N}\sum_{i=1}^{N} \mu_i \mu_i' \\ \hat{\sigma_\epsilon} &= \frac{1}{NT}||Y - X\beta - W\mu||^2. \end{aligned}$$

6.2 EM algorithm for longitudinal data

To find the maximum likelihood the EM algorithm can be use as describe by Laird and Ware [6].

$$Y_i = X_i\beta + W_i\mu_i + \epsilon_i.$$

Y_i is the vector of T outcomes on the ith individual. W_i is the design matrix for the individual random effect. μ_i is the individual random effect and ϵ_i is the vector $(\epsilon_{i1}, ..., \epsilon_{iT})'$. The parameter to estimate is $\theta = (\beta, R, \sigma_\epsilon^2)$

If μ is know the parameters to estimate are the following.

$$\hat{\beta} = (X'X)^{-1}X'(Y - W\mu)$$

$$\hat{R} = \frac{1}{N}\sum_{i=1}^{N}\mu_i\mu_i'$$

$$\hat{\sigma_\epsilon} = \frac{1}{NT}||Y - X\beta - W\mu||^2 = \frac{1}{NT}\big(||y - X\hat{\beta}||^2 + ||W\mu||^2 - 2 < y - X\beta, \hat{W}\mu > \big).$$

Or

$$\begin{aligned}
||W\mu||^2 &= \sum_{i=1}^{N}||W_i\mu_i|| \\
&= \sum_{i=1}^{N}\mu_i W_i'W_i\mu_i \\
&= \sum_{i=1}^{N}Trace(\mu_i W_i'W_i\mu_i) \\
&= \sum_{i=1}^{N}Trace(W_i'W_i\mu_i\mu_i').
\end{aligned}$$

So the statistics use to estimate θ are $\mu_1, ..., \mu_N$ and $\mu_1\mu_1, ..., \mu_N\mu_N$

Since μ and $\mu\mu'$ is not know the idea is to take there expectation.
The distribution for μ is:

$$\begin{aligned}
p(\mu_i|y_i;\theta) &\sim p(y_i|\mu_i;\theta)p(\mu_i;\theta) \\
&\sim C_1 \exp\big(-\frac{1}{2\sigma_\epsilon}||Y_i - X_i\beta - W_i\mu_i||^2 - \frac{1}{2}\mu_i'\sigma_\mu^{-1}\mu_i\big) \\
&\sim C_2 \exp\big(-\frac{1}{2}(\mu_i - \eta_i)'\tau_i(\mu_i - \eta_i)\big).
\end{aligned}$$

With

$$\tau_i = \big(\frac{W_i'W_i}{\sigma_\epsilon^2} + R_T^{-1}\big)^{-1} \quad ; \quad \eta_i = \frac{\tau_i W_i'(Y_i - X_i\beta)}{\sigma_\epsilon^2}.$$

So:

$$\begin{aligned} E[\mu_i|Y_i;\theta] &= \eta_i \\ E[\mu_i\mu_i'|Y_i;\theta] &= \mathrm{Var}(\mu_i|Y_i;\theta) + E[\mu_i|Y_i;\theta]E[\mu_i|Y_i;\theta]' \\ &= \tau_i + \eta_i\eta_i'. \end{aligned}$$

The estimated parameters become:

$$\hat{\beta} = (X'X)^{-1}X'(Y - WE[\mu|y,\theta])$$

$$\sigma_\epsilon^2 = \frac{1}{NT}\Big(||Y - X\hat{\beta}||^2 + \sum_{i=1}^{N}\mathrm{Trace}(W_i'W_i\mu_i\mu_i') - 2\sum_{i=1}^{N}(y_i - X_i\hat{\beta})'W_iE(\mu_i|y,\theta)\Big).$$

$$R = \frac{1}{N}E\big[\sum_{i=1}^{N}\mu_i\mu_i'|y_i,\hat{\theta}\big] = \frac{1}{N}\sum_{i=1}^{N}(\eta_i\eta_i' + \tau_i).$$

The formula $E[\mu_i|Y_i;\theta] = \eta_i$ give an empirical Bay estimator of μ_i. The algorithm can be resume as: For a θ_k estimate $E[\mu_i|Y_i;\theta_k]$ and $E[\mu_i\mu_i'|Y_i;\theta_k]$ and construct θ_{k+1} with the previous formula.

6.3 EM algorithm for mixed effects models with censored data

The EM algorithm for mixed effect model is here modify to deal with censored data. Hughes (1999) [7], [8].

$$Y_i = X_i\beta + W_i\mu_i + \epsilon_i.$$

The complete data is $(Y_i, \mu_i, \epsilon_i)_{i=1,..,m}$ the observed data is (C_i,Q_i).

$$\begin{cases} Y_{ij} = C_{ij} & \text{if } Q_{ij} = 0 \\ Yij > C_{ij} & \text{if } Q_{ij} = 1 \end{cases}$$

For censored data

$$\hat{\beta} = (X'X)^{-1}X'(E[Y|C,Q] - WE[\mu|C,Q])).$$

$$\hat{R} = \sum_{i=1}^{m} E(\mu_i\mu_i'|C_i,Q_i,\theta)/m.$$

$$\hat{\sigma_\epsilon}^2 = \frac{1}{NT}\Big(||E[Y|C,Q]-X\hat{\beta}||^2+\sum_{i=1}^{N}\text{Trace}(E(W_i'W_i\mu_i\mu_i'|C_i,Q_i,\theta))-2\sum_{i=1}^{N}(y_i-X_i\hat{\beta})'W_iE(\mu_i|C_i,Q_i,\theta)\Big).$$

To handle the censoring we integrate over the censored value.

$$E(\mu_i\mu_i'|C_i,Q_i,\theta) = \int_{Y_i(C,Q)} E(\mu_i\mu_i'|Y_i,\theta)f(Y_i|C_i,Q_i,\theta)$$

$$E(\mu_i|C_i,Q_i,\theta) = \int_{Y_i(C,Q)} E(\mu_i|Y_i,\theta)f(Y_i|C_i,Q_i,\theta)$$

With $f(Y_i|C_i,Q_i,\theta)$ the truncated multivariate normal with mean $X\beta+W\mu$, variance σ_ϵ^2 and left limit l. $E(\mu_i\mu_i'|Y_i,\theta)$ and $E(\mu_i|Y_i,\theta)$ are the expected complete data sufficient statistics.

$$\begin{aligned} E[\mu_i|Y_i;\theta] &= \eta_i \\ E[\mu_i\mu_i'|Y_i;\theta] &= \text{Var}(\mu_i|Y_i;\theta) + E[\mu_i|Y_i;\theta]E[\mu_i|Y_i;\theta]' \\ &= \tau_i + \eta_i\eta_i'. \end{aligned}$$

With:

$$\tau_i = \big(\frac{W_i'W_i}{\sigma_\epsilon} + R_T^{-1}\big)^{-1} \quad ; \quad \eta_i = \frac{\tau_i W_i'(Y_i - X_i\beta)}{\sigma_\epsilon}.$$

The integral is calculated with Monte Carl integration.

$$E(\mu_i\mu_i'|C_i, Q_i, \theta) = \sum_{l=1}^{L} E(\mu_i\mu_i'|Y_i^l, \theta)/L$$

$$E(\mu_i|C_i, Q_i, \theta) = \sum_{l=1}^{L} E(\mu_i|Y_i^l, \theta)/L$$

$$E(Y_i|C_i, Q_i, \theta) = \sum_{l=1}^{L} Y_i^l/N$$

With Y_i^l following the truncated normal distribution on the censored variable and a degenerate distribution on the other. And L is the number of Monte Carlo sample

An asymptotic approximation for the variance of the fixed effects is given by:

$$\mathrm{Var}(\beta) = \Big(\sum_{i=1}^{N} X_i'Z_iX_i - X_i'Z_iB_iZ_iX_i\Big)^{-1}$$

With $B_i = \mathrm{Var}(Y_i|C_i, Q_i, \theta)$

One of the main problems of this algorithm is that it is computationally heavy. Indeed, it is necessary to compute $E[\mu_i|Y_i, \theta]$ and $E[\mu_i|Y_i, \theta]$, then $E(\mu_i|C_i, Q_i, \theta)$ and $E(\mu_i|C_i'|C_i, Q_i, \theta)$. The calculation can be simplified by using a Gibbs sampler. At each iteration, we sample $y_i \sim p_{yi}(y_i|b_i, C_i, Q_i, \theta)$ and $\mu_i \sim p_{\mu i}(\mu_i|y_i, Q_i, C_i) = p_{\mu i}(\mu_i|y_i, \theta)$ with the same distribution and parameters as in the study of the Gibbs sampler in the previous part.

7 Residual Analysis

To justify the use of the Tobit linear mixed model a analyse of the residual is made. Three type of residual can be studied with the linear mixed model .

The marginal residual $\hat{\xi} = Y - X\hat{\beta}$, that predict the marginal errors ξ=Y-E[Y]=Y-Xβ

The conditional residual $\hat{\epsilon}$=Y-$X\hat{\beta}$-W$\hat{\mu}$, that predict the conditional errors ϵ=Y-E[Y—μ]=Y-Xβ-Wμ

The BLUP W$\hat{\mu}$, that predict the random effect,Wμ=E[Y—μ]-E[Y]

The residual analysis is done with the EM algorithm. The Gibbs sampling give similar result.

7.1 Gibbs

7.1.1 Marginal residual

The variance of Y_i can be estimated by $\xi\xi'$. To study the within-subjects covariance matrix, we can use the Frobenius norm of $V_i - \xi\xi'$ (with V_i the variance of Y_i), $||V_i - \xi\xi'||^2$ must be close to zero. The following graph is the plot of this value as a function of the subjects' indices.

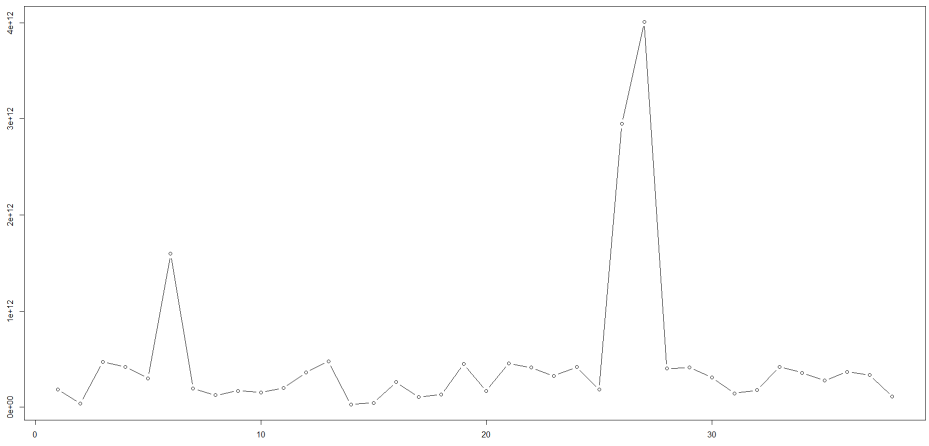

Figure 7: Subject indices versus $||V_i - \xi\xi'||^2$

The assumed covariance structure does not fit well in at least two cases. (for subjects 26 and 27 and possibly 6).

7.1.2 Conditional residual

The conditional residual $\hat{\epsilon} = X\hat{\beta} + W\hat{\mu}$, can be used to asses the presence of outlying observation, the homoscedesticity of the conditional errors and his normality.

The next plots is the conditional residual over the fitted value.

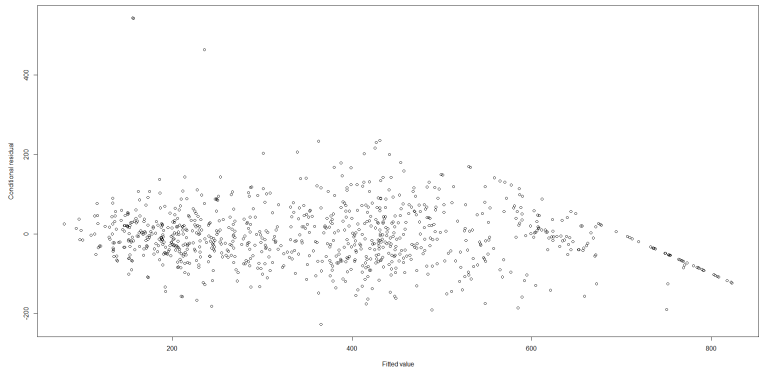

Figure 8: Fitted value versus Conditional residual

The censored value still cause the clear line. The line correspond to 700 minus the fitted value. Despite the patern the homoscedesticity seem respected.

The presence of outlying observation can be detected with the plot of the standardized conditional residual versus the observation indices

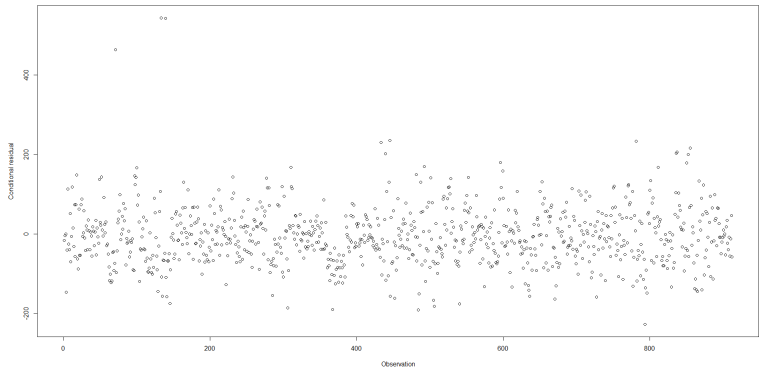

Figure 9: Observation indices versus conditional residual

There is 3 outlying observation at 71,133 and 139.

The normality of the conditional errors can be verified with a QQ plot of the standardize conditional residual.

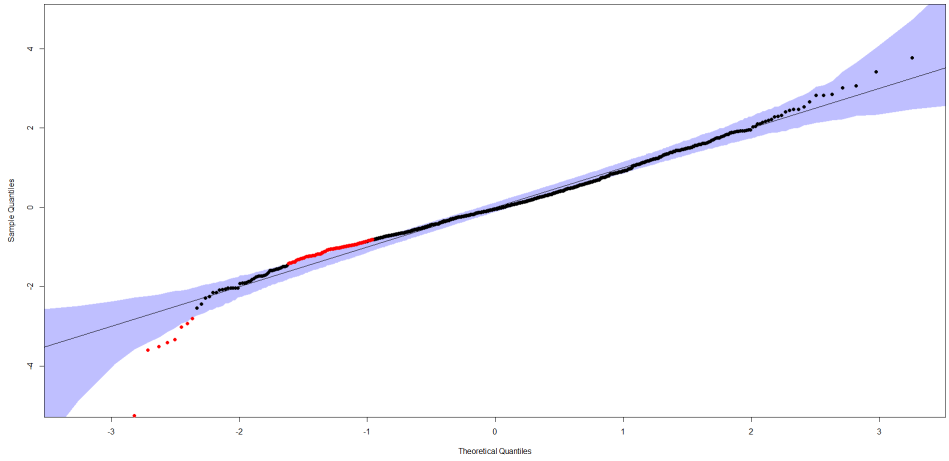

Figure 10: QQ plot

All the points are not in the 95% tolerance band but it is better than the plot without taking into account censoring.

7.1.3 Best linear unbiased predictions (BLUP)

$W\hat{\mu}$ predict the random effects, $W\mu = E[Y|\mu] - E[Y]$. So $W_i\mu_i$ reflects the difference between the predicted responses for the i-th subject and the population average; therefore it can also be used to find outlying subject and verified the normality of the random effects.

The two next plot is the EBLUP versus the subject indices.

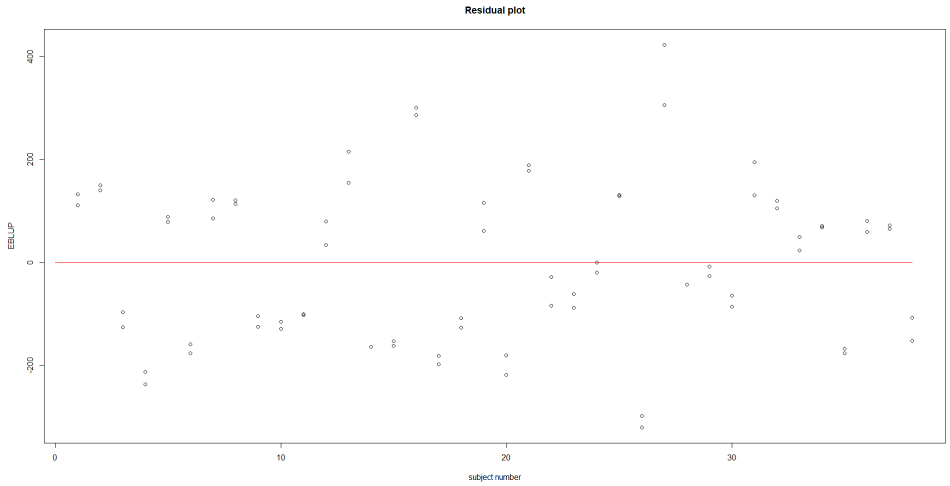

Figure 11: Subject indices versus BLUP

Subject 27 is different from the others

The next plot is the random effect versus the subject number. The graph plot the random intercept and the side random effect (left to rigth)

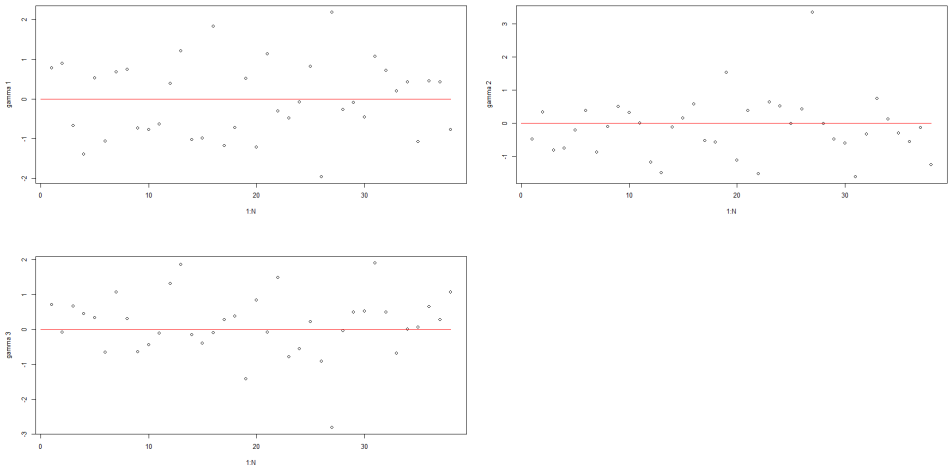

Figure 12: Subject indices versus the 3 random effect

Here again the subject 27 is different from the others over.

The subject 27 have 13 censored measurement. It is normal that it does not respect the model since many of its values are constant and equal to 700.

The other all study of residual doesn't show any strong contradiction to the model. Most of the strange behavior is due to the censored point, which is actually taken into account in the algorithms. We can considered this model as valid for this dataset.

8 Analyse of the results

8.1 Gibbs Sampling

In the section we focus on the result give by the Gibbs sampler. The first plot is 10000 iteration of the sampler:

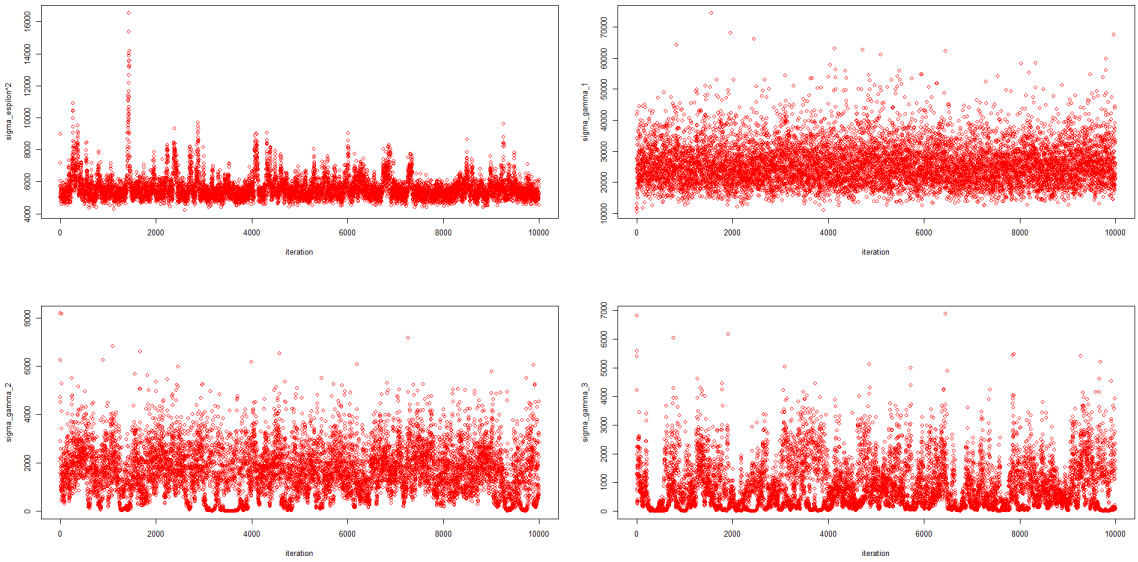

Figure 13: 10000 iteration of the Gibbs sampling for σ_ϵ,$\sigma_{\gamma 1}$,$\sigma_{\gamma 2}$,$\sigma_{\gamma 3}$ from left to right

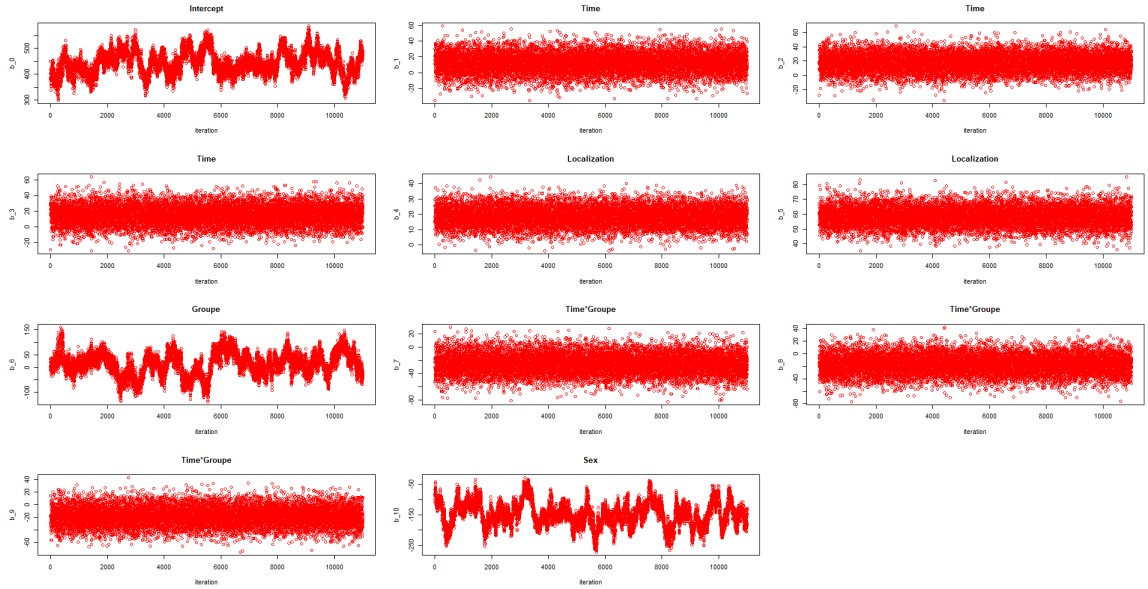

Figure 14: 10000 iteration of the Gibbs sampling β parameters

The intercept,group and sex are slower to converge. It show strong autocorelation for this parameters. The other group seem to converge very quickly to there distribution.

The next graph show the convergence of five Gibbs sampling on the different initial value with a burnin of 1000 iteration.

As show the Gibbs sampling is stable regarding the choice of the initial value. Even with initial value far from coherent value the Gibbs sampling converge . The blue line is the following value:

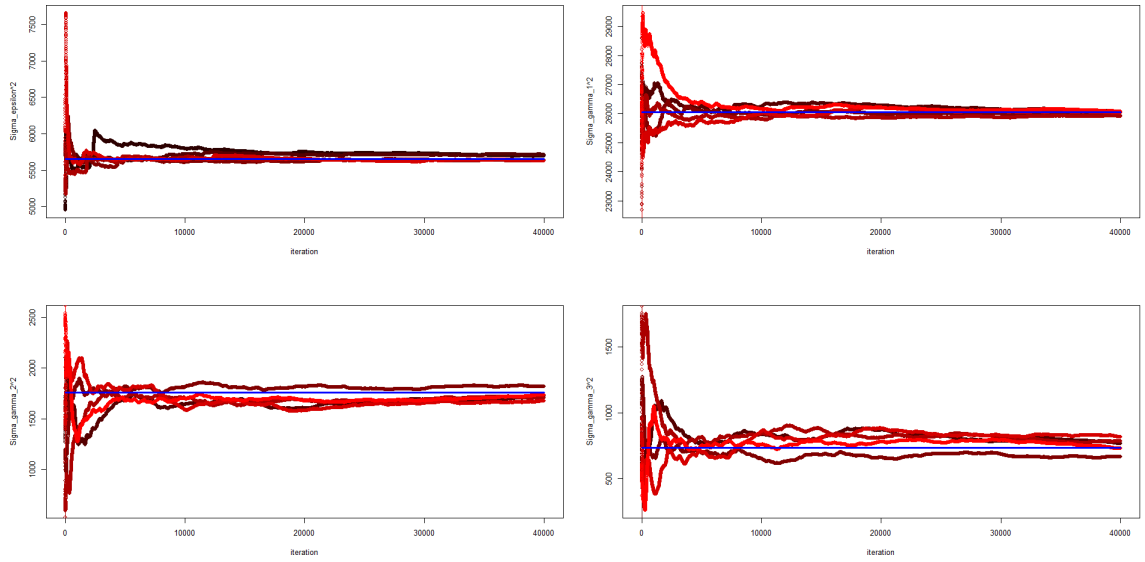

Figure 15: Convergence of the variance for different initial value

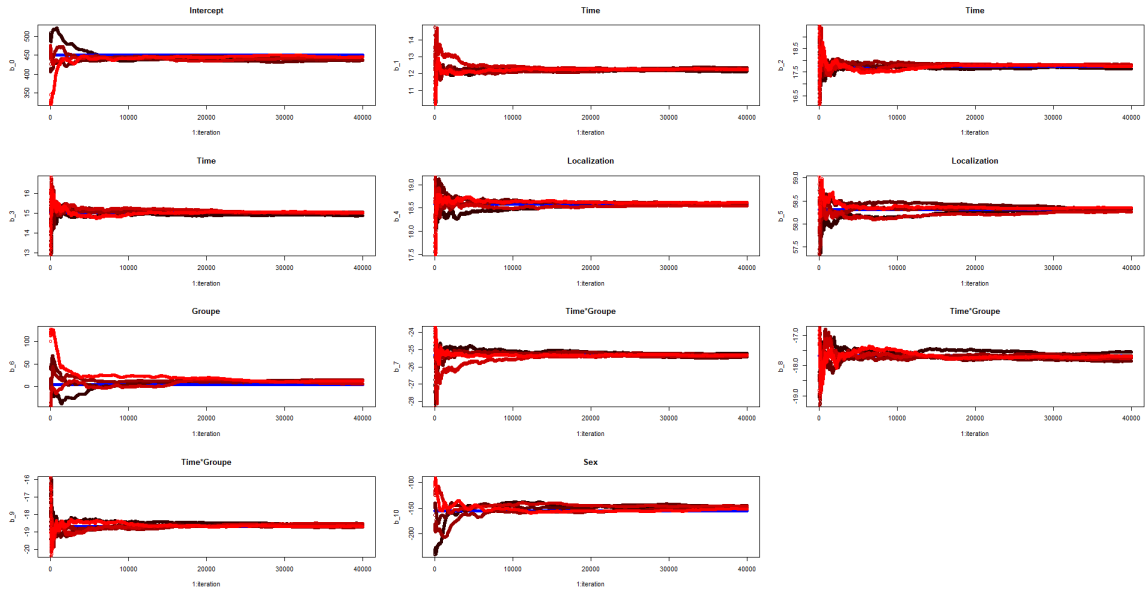

Figure 16: Convergence of β for different initial value

		Value	Standard error
Variance error	σ^2_ϵ	5652.47	4.16
Variance random effect	$\sigma_{\gamma 1}$	26053.24	33.26
Variance right	$\sigma_{\gamma 2}$	1762.03	5.19
Variance left	$\sigma_{\gamma 3}$	733.9594	3.95
Intercept	$\beta 0$	441.89	0.20
Time 2	$\beta 1$	12.23	0.06
Time 3	$\beta 2$	17.73	0.06
Time 4	$\beta 3$	15.03	0.06
Trapezius 2	$\beta 4$	18.58	0.03
Trapezius 3	$\beta 5$	58.31	0.03
Group	$\beta 6$	16.30	0.22
Time2× Group	$\beta 7$	-25.33	0.07
Time3× Group	$\beta 8$	-17.67	0.07
Time4× Group	$\beta 9$	-18.71	0.07
Sex	$\beta 10$	-154.97	0.20

The variance of the random error is greater than that of the random effect, which is consistent. The random difference between patients is the main random effect. The variance of the left and right is less than the variance of the random error, which means that the side effect has no significant impact. In this study, we are interested in the variation over time. The parameters of interest are Time and Time×Group. By adding the time and the time group for the patient, we can see the variation of the pain threshold. This variation pattern is illustrated in the following graphs.

The 6 regression mean for the patient follow the same pattern. A decrease in pain thresholds after the exercise and then the value returns to the starting value and finally a small decrease for the last value. This pattern was already detected with the mean of the measurement but it was clear only on the right side.

The following plot is the regress value for the two group. The control group is in red.

Most regression values follow the same pattern of variation. Only the regression value for two individuals shows a different pattern of variation.

The figure 18 represents the average pain regression threshold. The control group has a higher pain threshold, which means that they are more resistant to pain, as expected. For the patient, the pain threshold decreases on the second measurement, i.e. after the exercise. This shows a higher pain sensitivity for the patient, unlike the control group, which has a higher pain threshold than the initial measurement. The patient returns to his initial pain threshold after one hour. After exercise, the control group maintained the high pain thresholds.

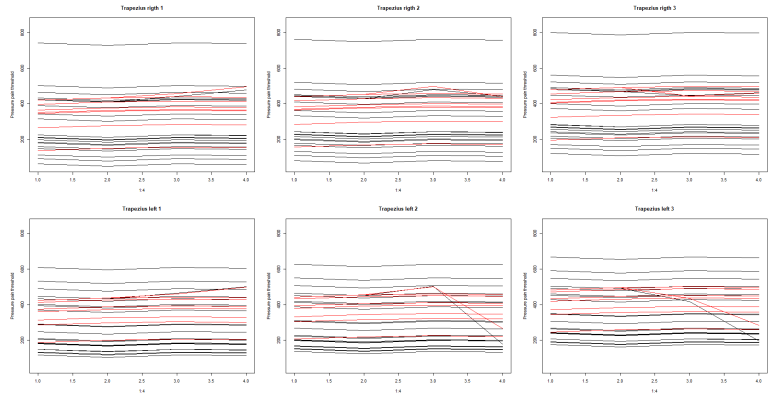

Figure 17: Regression value. The top 3 graphs are the 3 trapezius on the right side and the bottom 3 graphs on the left side.

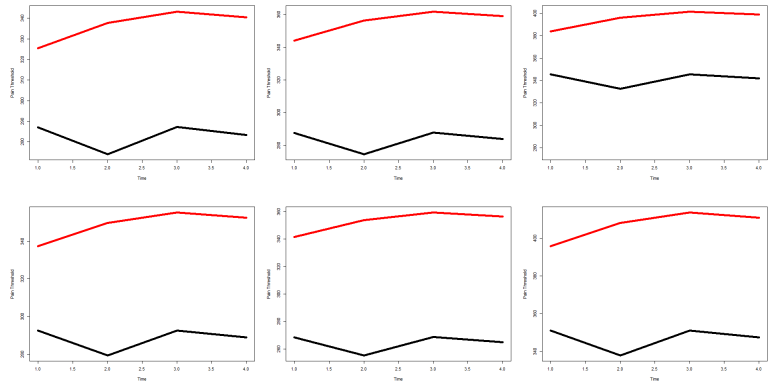

Figure 18: Average of the regression values for the 3 trapezius on the right (top) and left (bottom)

The post-exercise decrease for the patient is -13.1. Since we are looking into the variation of the point only the variance of the residuals should be look at. The error variance is σ_ϵ^2=5652 and there are 26 patients. It lead to a standard error of 14,7 for the mean of the patient. It lead to a 75% confidence interval of [-23,127;-3,073] using the student distribution. In comparison of the 12.23 increase of the control group with a standard error of 21,703 which lead to a 75% confidence interval of [-2,526;26,98]. This interval lead to the conclusion that a light physical exercise lead to more pain sensitivity for the patient.

The last measurement (the next day) shows a decrease of -3.68 for the patient and an increase of 15 for the control group compared to the initial value. Which lead to a 75% confidence interval of [-13,70;6,347] for the patient and [0,742;30,25] for the control. The interval are less statically significant but exercise tends to reduce or at least not increase the pain threshold for the patient in contrast to the control wish which tends to have a higher pain threshold. The finding for the change after one day after exercise is less statically significant.

8.2 The EM algorithm

The result of the EM algorithm is here study. The first graph is 100 iteration of The EM algorithm with the censored data.

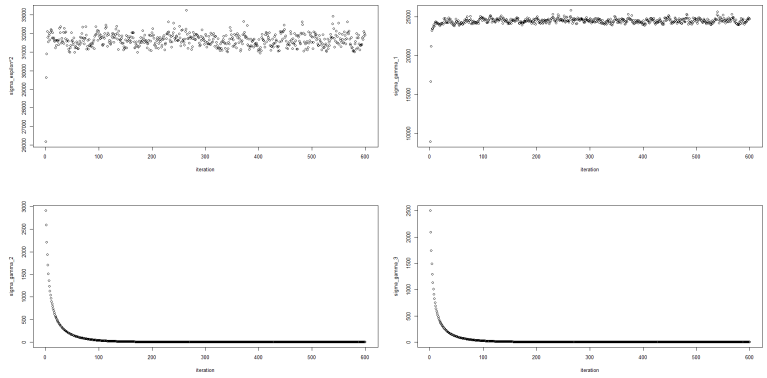

Figure 19: 100 iteration of the EM algorithm. Variance parameters

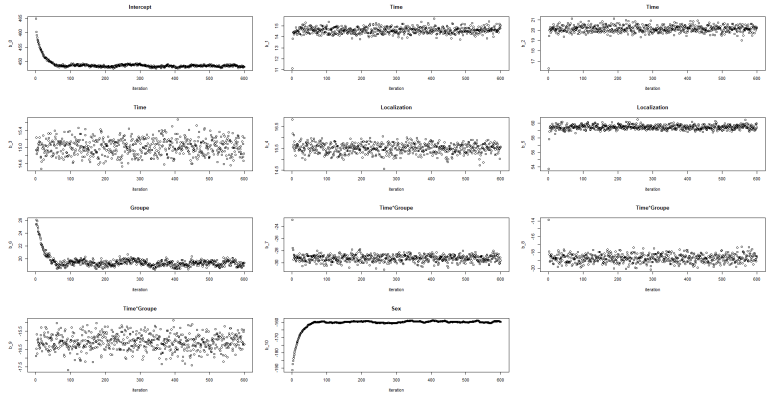

Figure 20: 100 iteration of the EM algorithm. β parameters

The EM algorithm converge in about 100 iteration. The Monte Carlo integration cause small variation.The most interesting parameters (time and time group) converge quickly and have a very low variation.

The next graph show the convergence of 5 EM algorithm on different initial value. For $\sigma_{\gamma j}$ in (1,100,1000,10000,100000), σ_ϵ in (1,100,10000,10000,100000). β is initialise using the maximum likelihood formula on the initial censored dataset and $\mu \sim \mathrm{N}(0, \sigma_{\gamma j})$. As shown in the figure 21 the algorithm converge even with initial value far from the true value but take much more iteration to converge.

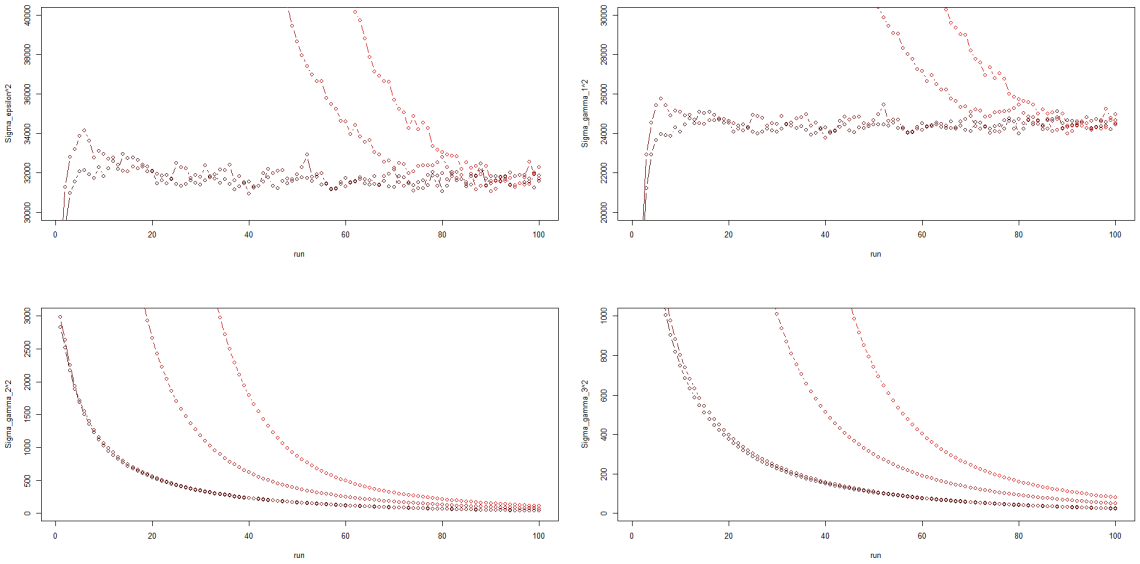

Figure 21: Variances parameters on different initial value

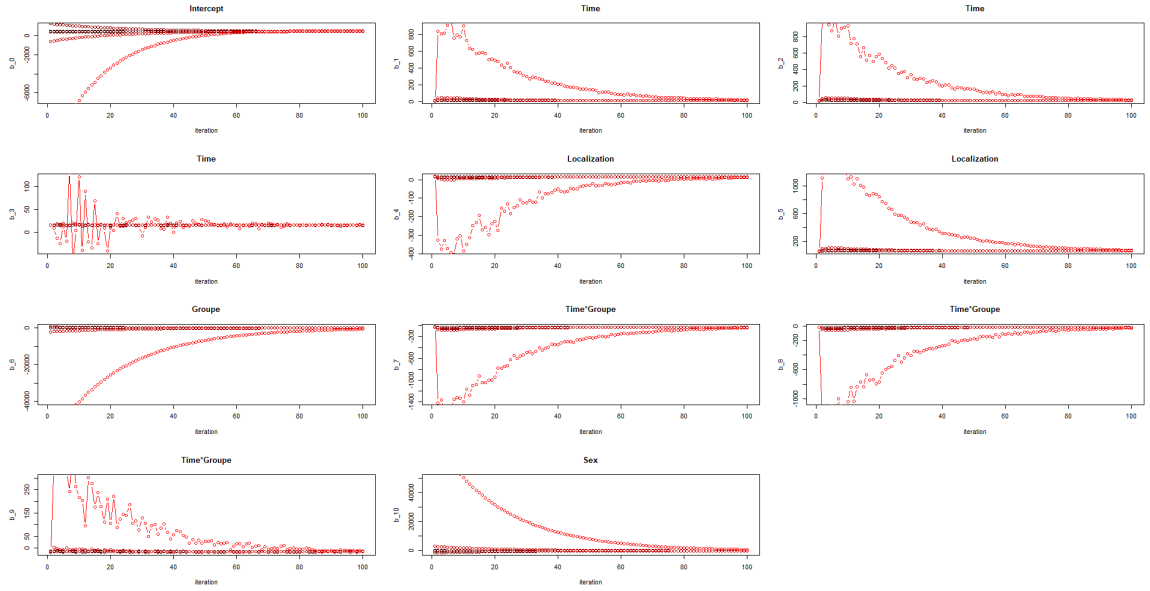

Figure 22: β parameters on different initial value

	Value	Standard error
$\beta 0$	436.907301	3.8
$\beta 1$	11.572917	0.99
$\beta 2$	16.767362	0.99
$\beta 3$	14.930556	0.99
$\beta 4$	16.678620	0.91
$\beta 5$	53.712205	0.91
$\beta 6$	2.229326	4.6
$\beta 4$	-23.330931	1.2
$\beta 8$	-14.482647	1.2
$\beta 9$	-16.615716	1.2
$\beta 10$	-135.949018	4.4
σ_ϵ^2	6159.873	
$\sigma_{\gamma 1}^2$	19681.28	
$\sigma_{\gamma 1}^2$	925.0956	
$\sigma_{\gamma 1}^2$	49.10442	

As the table shows, the variance of the random intercept is much lower than the previous algorithm. The side effect also tends to disappear. The EM algorithm give similar result to the Gibbs sampling.

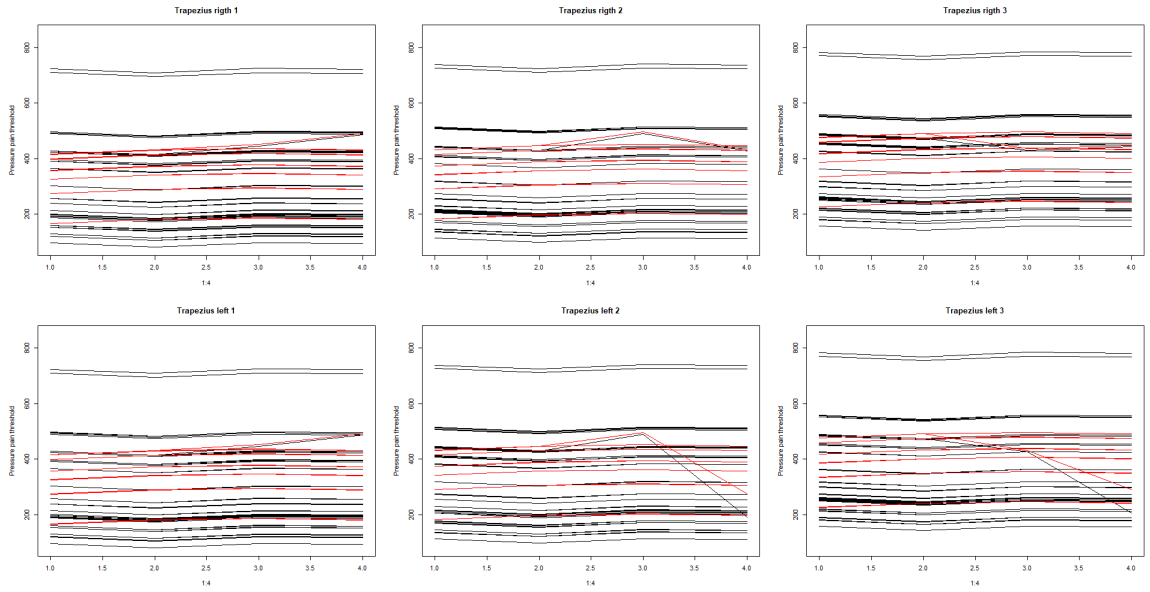

Figure 23: Regress data of the EM algorithm the control group in red

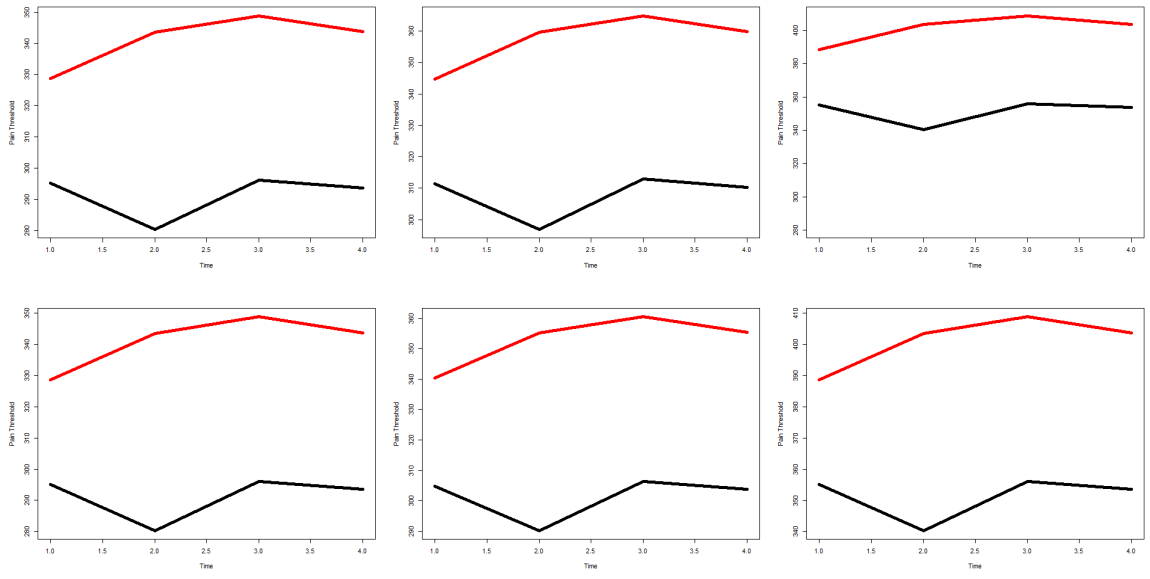

Figure 24: Mean value of pain threshold. The patient in black and control group in red

9 The impact of censoring

With the result of the algorithms on the experimental data, simulated data can be created. The idea is to evaluate the stability with different level of censoring of the algorithm on this simulated data set. As the actual value is known, the efficiency of the algorithm can be easily studied. The variation over time is studied, so the important parameters to study are time and the interaction between time and group. The other variables will only result in a translation of the value of the 4 time point.This translation is only important to study the difference between the groups ,sex and localisation.

9.1 Gibbs sampling

9.1.1 On the regressed value

We first create the simulated data as $Y_{sim} = X\hat{\beta} + W\hat{\mu}$.

The intercept and the variance is resume in the following table.

	True value	0%	5%	13%	23%	40%
without intercept	437	427	442	445	409	351
with intercept	437	431	512	418	427	417
without σ_ϵ^2	5664	1896	2002	2066	1773	1190
with σ_ϵ^2	5664	1841	0	0	0	0
without $\sigma_{\gamma 1}^2$	26071	26892	25858	21915	16777	9893
with $\sigma_{\gamma 1}^2$	26071	26542	28219	26053	30354	35423
without $\sigma_{\gamma 2}^2$	1741	939	608	312	315	294
with $\sigma_{\gamma 2}^2$	1741	1000	1931	2472	1001	4499
without $\sigma_{\gamma 3}^2$t	726	474	379	347	237	90
with $\sigma_{\gamma 1}^3$	726	418	479	409	855	2174

The variance of the random error drop to zero when censoring is taken into account. As shown in the next graph the Gibbs sampling with the truncated normal distribution estimated very well the Time and Time×Group the parameters.

The figure 25 is the sum of the β parameters time and the interaction for time and group for different level of censoring. It is this sum who will direct the variation over time for the patient. The first time is coded as 0 so the value are the variation from the initial value. The true value is in blue. The more censored the data is the more red are the curve. Almost only one curve can be seen. The algorithm estimates these parameters perfectly even with a high level of censoring. The first value in the graph is negative which means that there is a decrease in the pain threshold for the second measurement as expected. The figure 26 is without taking into account censoring. The impact of censoring can be clearly seen. The first and last values tend to increase with increasing level of censoring, which has the effect of flattening the curve of the average pain threshold.

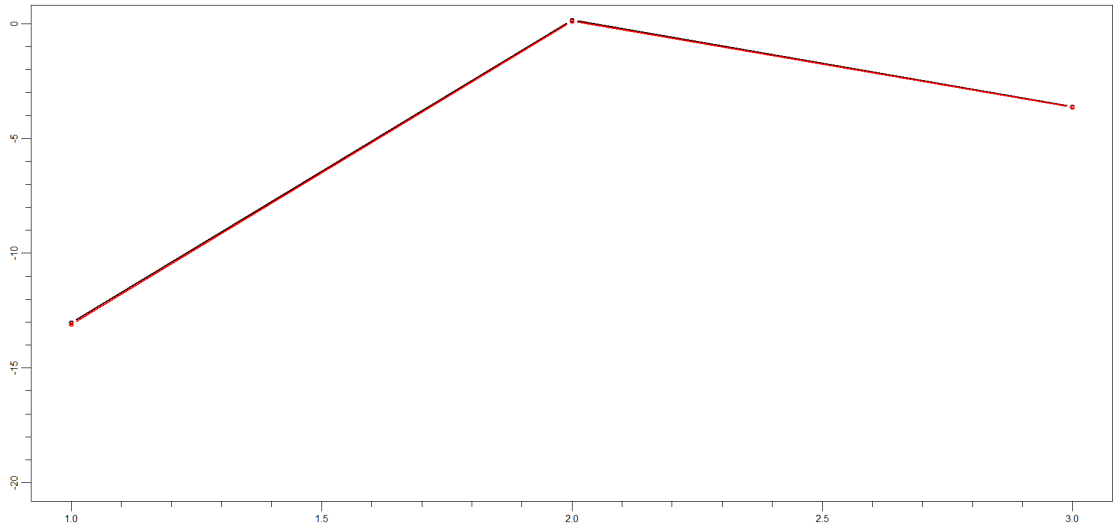

Figure 25: Time+Time×Group with censoring taking into account

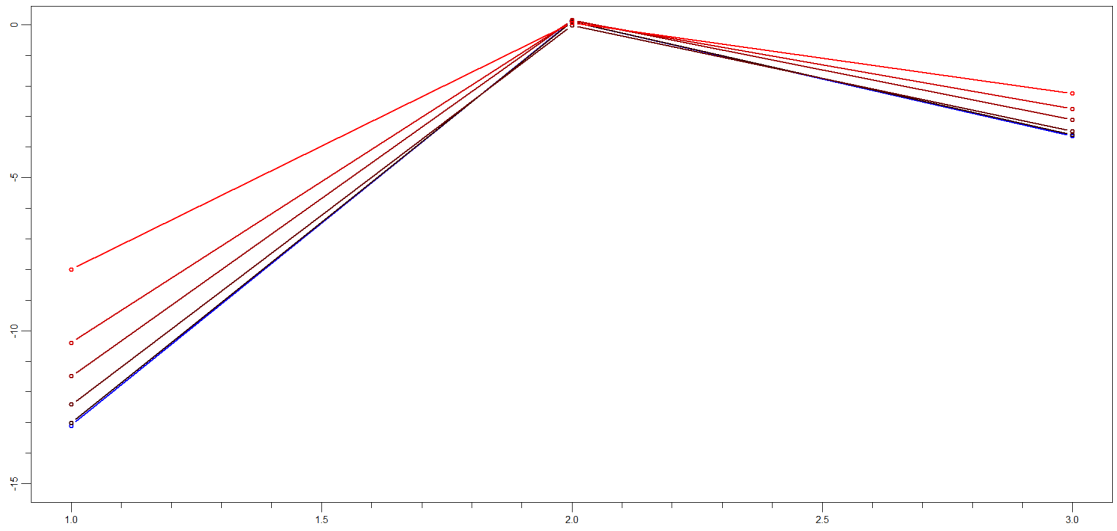

Figure 26: Time+Time× Group parameter without censoring taking into account

Figures 26 and 27 show the time parameters with the truncated normal distribution and the second without. These parameters are responsible for the variation in the control group. All 3 values are positive, which means an increase in pain threshold after exercises. The same conclusions as for the patient are shown.

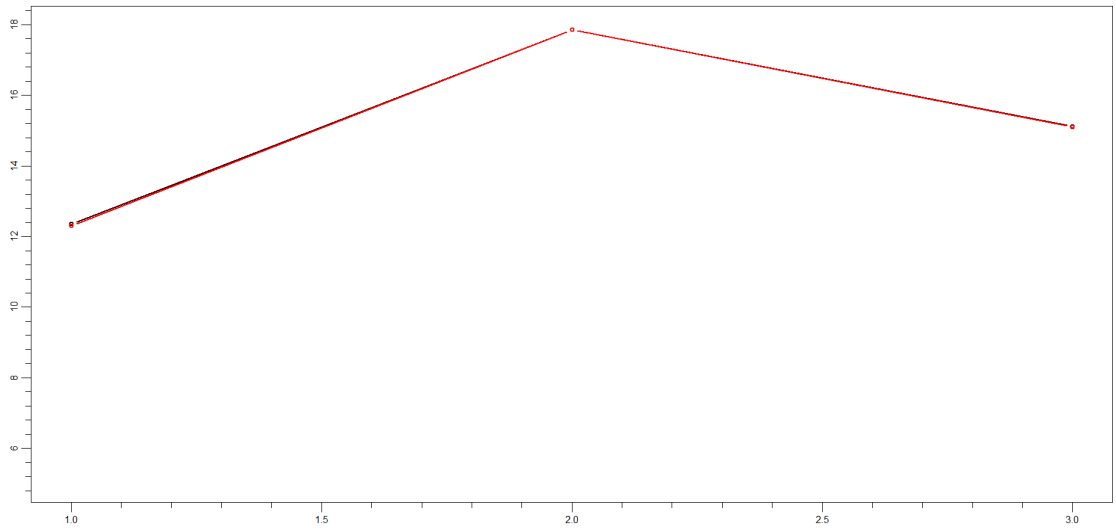

Figure 27: Time parameter with censoring taking into account

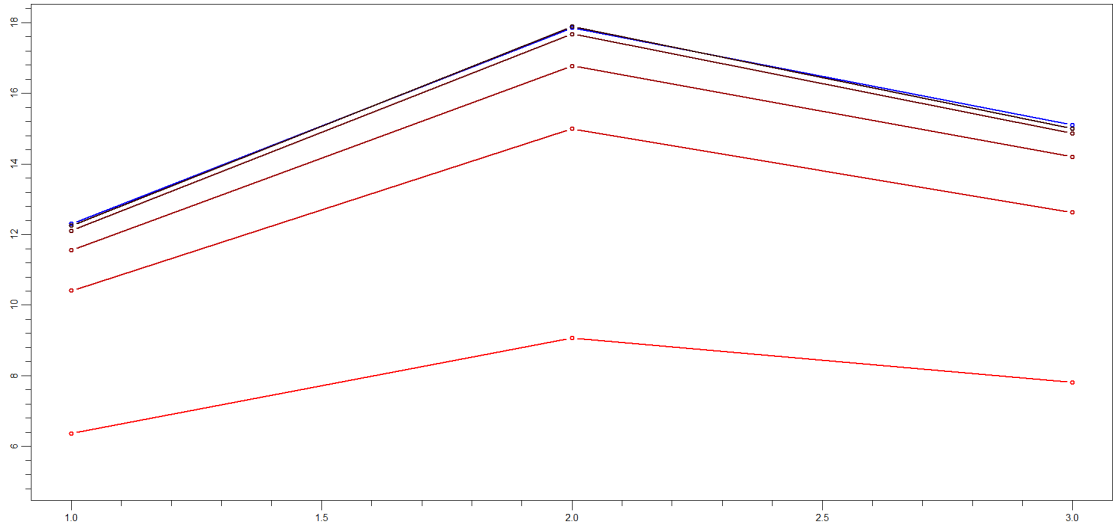

Figure 28: Time parameter without censoring taking into account

The variation pattern can be seen in the mean plot of the regress value. The variation is as show in the previous graph almost the same for all the level of censoring. A height level of censoring lead to a small overestimation of the value of the mean if the truncated normal distribution is used. Not taking account of censoring lead to underestimation of the value of the mean and tend to fade the variation. The same conclusion can be seen for the control group.

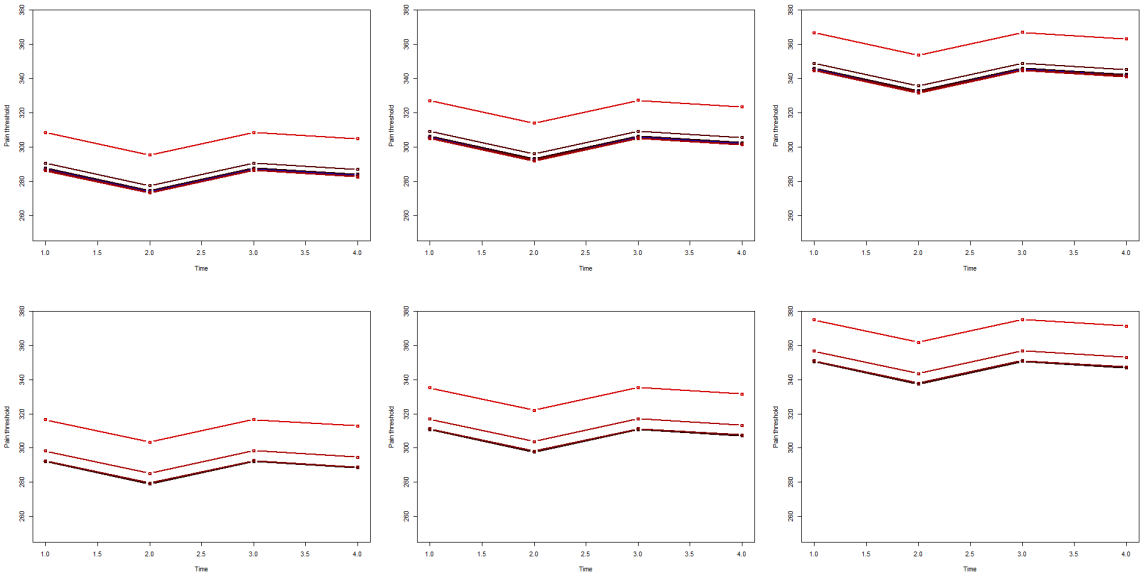

Figure 29: Mean of the patient with censoring taking into account

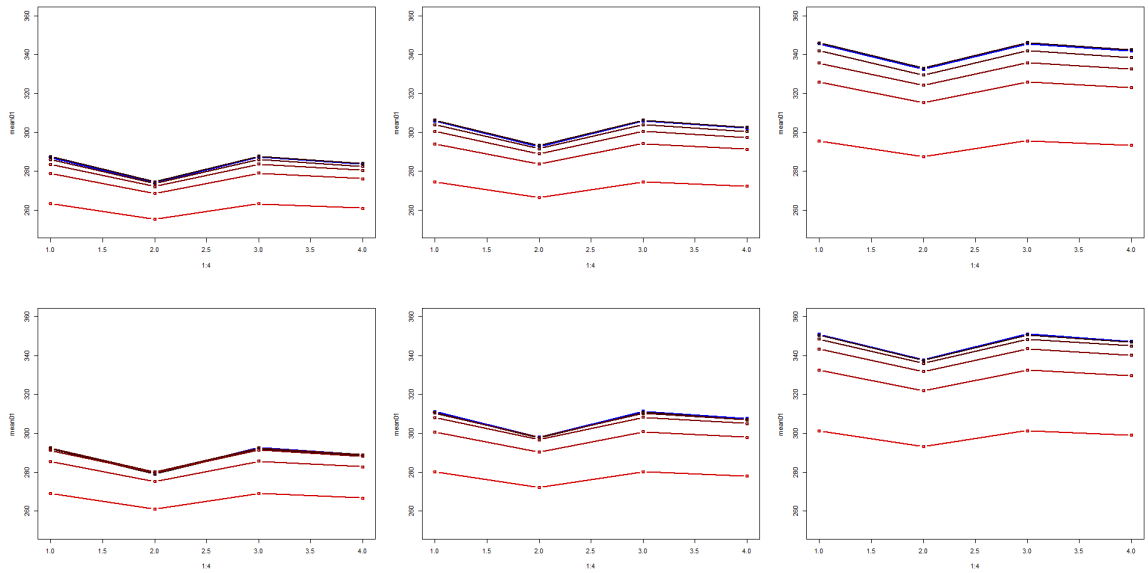

Figure 30: Mean of the patient without censoring taking into account

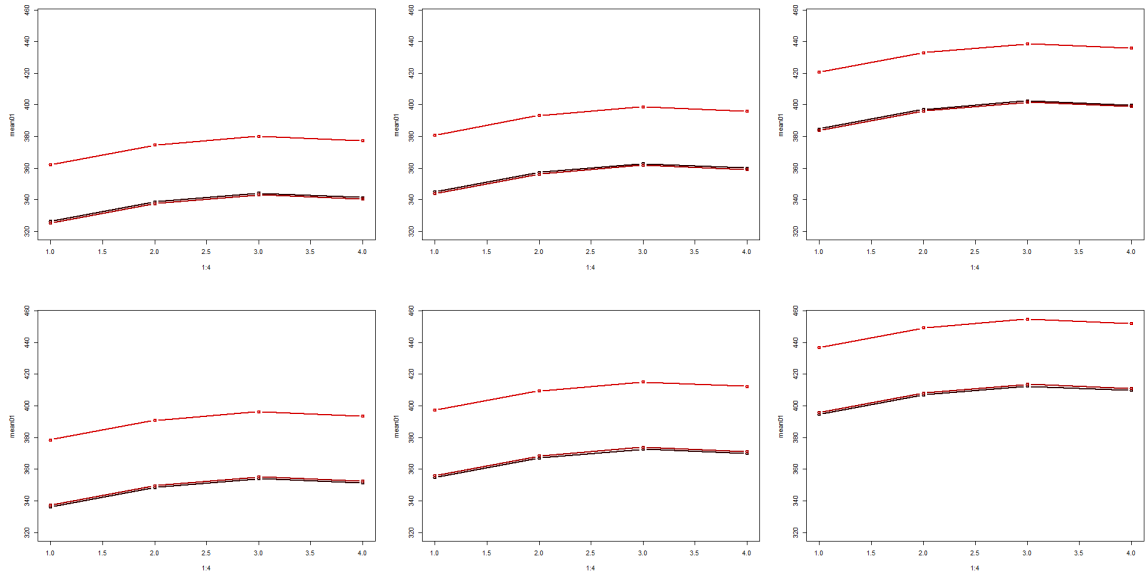

Figure 31: Mean pain threshold of the control group with censoring taking into account

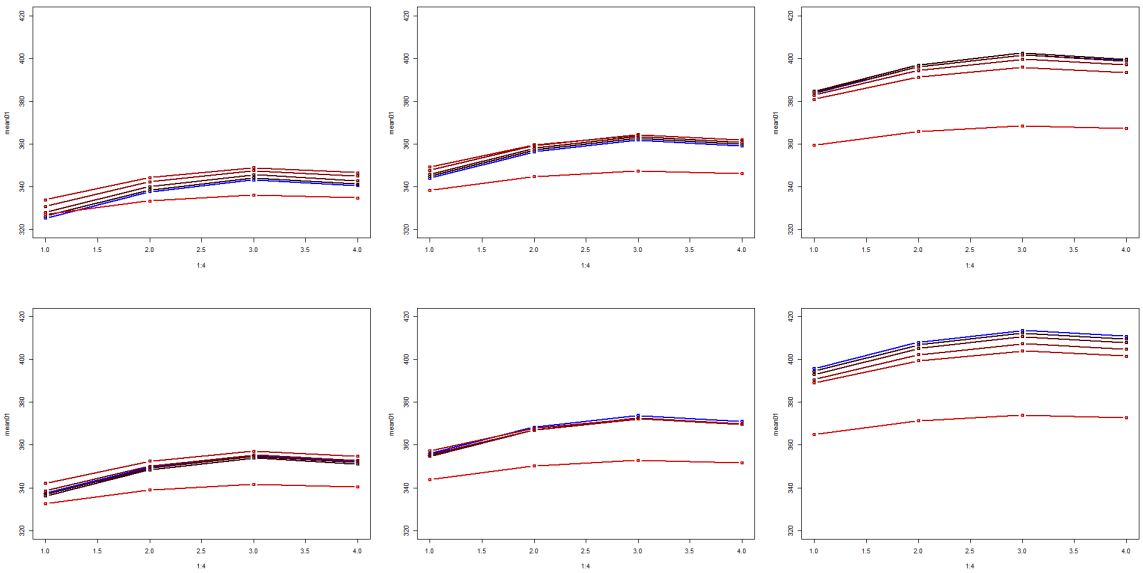

Figure 32: Mean pain threshold of the control group without censoring into account

The Gibbs sampler that accounts for censoring works best with censored data. It tends to overestimate the data with a high level of censoring, but the variation over time remains the same. Gibbs sampling without using the truncated normal distribution tends to attenuate the variation and underestimate the mean value. The difference start to be significantly especially for more than 20% of censoring.

9.1.2 On regressed value with random error

In this part we simulated the data according to : $Y_{sim} = X\hat{\beta} + W\hat{\mu} + \epsilon$ with $\epsilon \sim N(0, \hat{\sigma_\epsilon}^2)$.

The following graph is the sum of the time and time-group interaction without censoring with the data for 5 different iterations of the algorithm. As can be seen, the impact of the random error is significant. With this data set, the random error is high compared to the impact of the censoring. The main focus is the difference between taking into account or not the censoring.

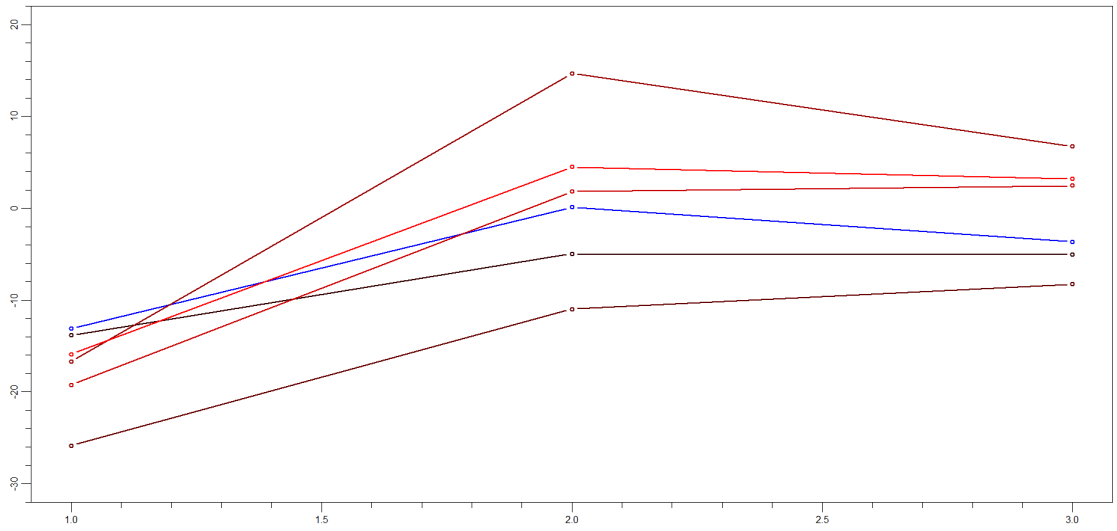

Figure 33: Times+Time×Group for 5 Gibbs sampling on data set with different random error

The following table resume the intercept and the variance component for different level of censoring.

	True value	0%	5%	13%	23%	40%
with intercept	437	431	435	439	434	469
without intercept	437	436	427	420	386	342
with σ^2_ϵ	5664	8916	6859	6450	6231	6037
without σ^2_ϵ	5664	9611	8966	8114	6494	4475
with $\sigma_{\gamma_1}^2$	26071	26776	26205	27171	24777	30135
without $\sigma_{\gamma_1}^2$	26071	26745	24742	21361.75	16148	9449
with $\sigma_{\gamma_2}^2$	1741	476	589	478	410.20	410.20
without $\sigma_{\gamma_2}^2$	1741	455	338	235	161	133
with $\sigma_{\gamma_3}^2$	726	412	578	477	438.49	305
without $\sigma_{\gamma_3}^2$	726	361	291	235	146.99	120

The intercept tends to decrease if censoring is not taken into account, conversely, if censoring is taken into account, the intercept is stable, especially for 23% and below. The random intercept also tends to decrease if censoring is not taken into account, which can be a problem for the calculation of the confidence interval. The random error is also better estimated with censoring taken into account.

The figure 34 and 35 is the graph of the Time+TimeGroup wish is responsible for the variation of the patient group. The first value remains negative, which means that the decrease in pain thresholds after exercise . The difference between the taking or not censoring is less visible.

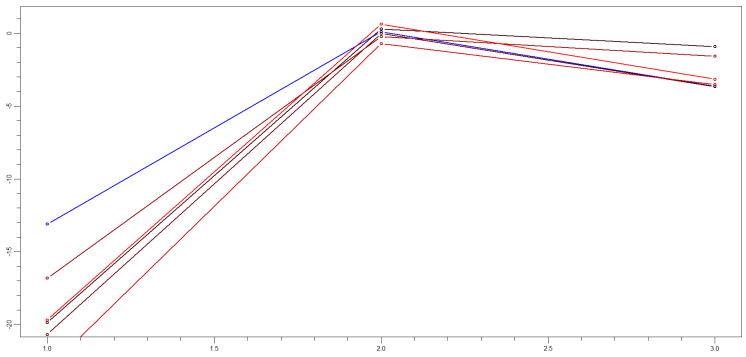

Figure 34: Time+Time×Group with censoring taking into account for 5 different level of censoring, the more red is the curve the more censored is the data

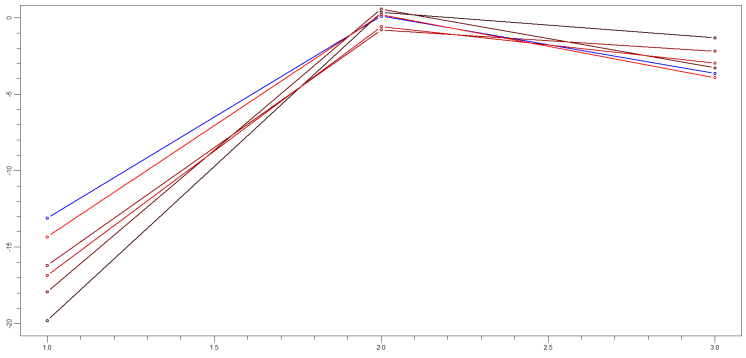

Figure 35: Time+Time×Group without censoring taking into account for 5 different level of censoring, the more red is the curve the more censored is the data

The graph 36 and 37 is the one of the Time parameter responsible for the variation of the control group.

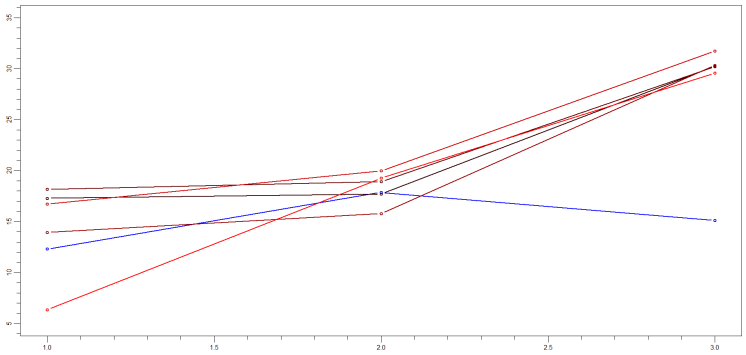

Figure 36: Times with censoring taking into account for 5 different level of censoring, the more red is the curve the more censored is the data

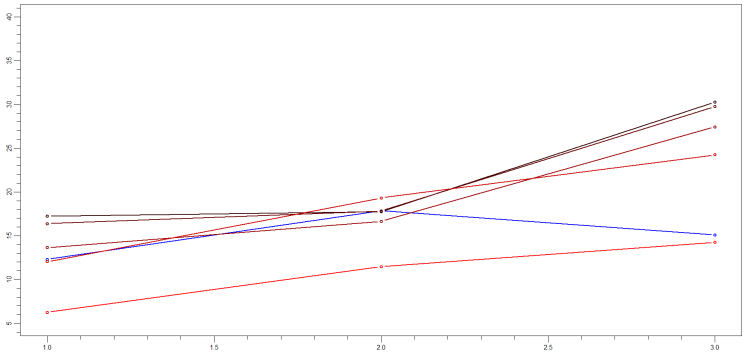

Figure 37: Times without censoring taking into account for 5 different level of censoring, the more red is the curve the more censored is the data

The mean are more stable with increasing level of censoring if the censoring is taken into account.

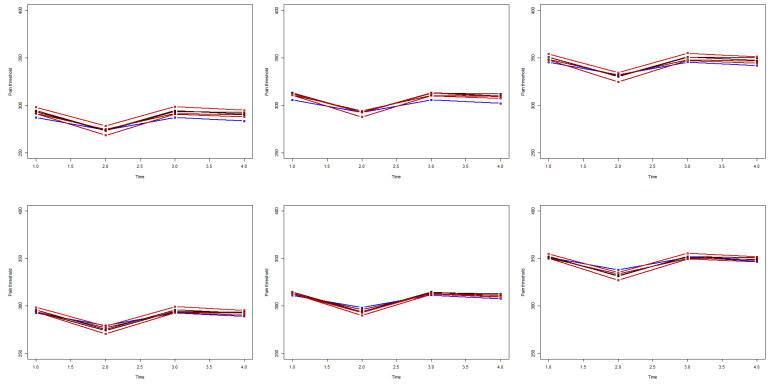

Figure 38: Mean of pain threshold for the patient with censoring taking into account for 5 different level of censoring, the more red is the curve the more censored is the data

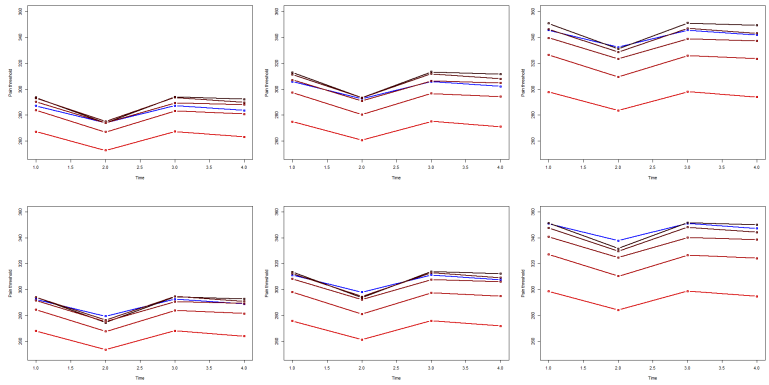

Figure 39: Mean of pain threshold for the patient without censoring taking into account for 5 different level of censoring, the more red is the curve the more censored is the data

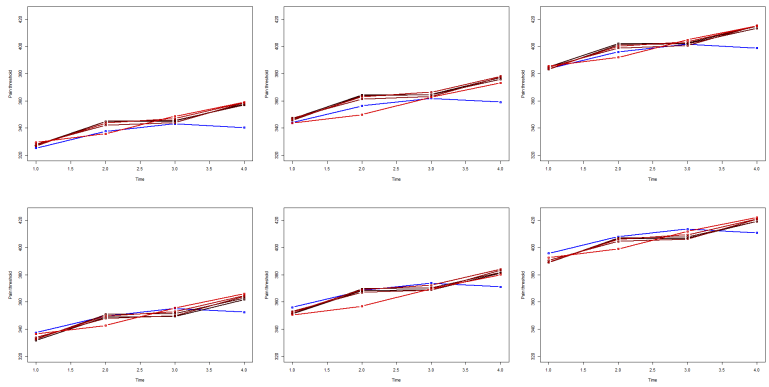

Figure 40: Mean of pain threshold for the control group with censoring taking into account for 5 different level of censoring, the more red is the curve the more censored is the data

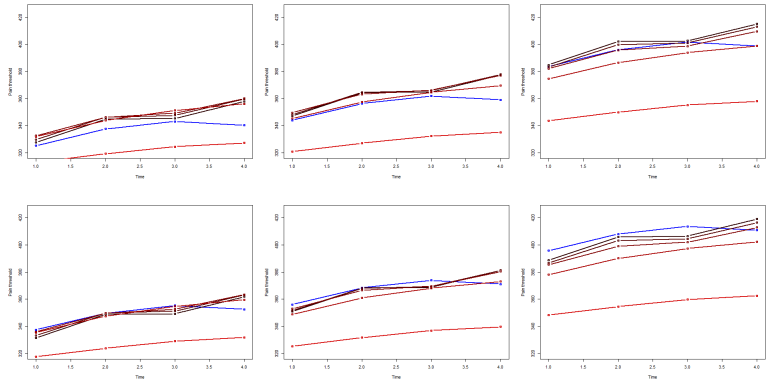

Figure 41: Mean of pain threshold for the control group without censoring taking into account for 5 different level of censoring, the more red is the curve the more censored is the data

Taking censoring into account led to a more stable result. It also leads to a better estimation of the average pain threshold with a slight overestimation. Conversely, not taking censoring into account leads to an underestimation and a less stable result.

The the mean squared error given by:

$$MSE = \frac{1}{N * T} \sum_{i=1}^{N*T} (Y_i - \hat{Y}_i)^2$$

With $\hat{Y}_i = X\hat{\beta} + W\hat{\mu}$ can be used to compare the efficiency.

The red line is the Gibbs sampling taking into account censoring.

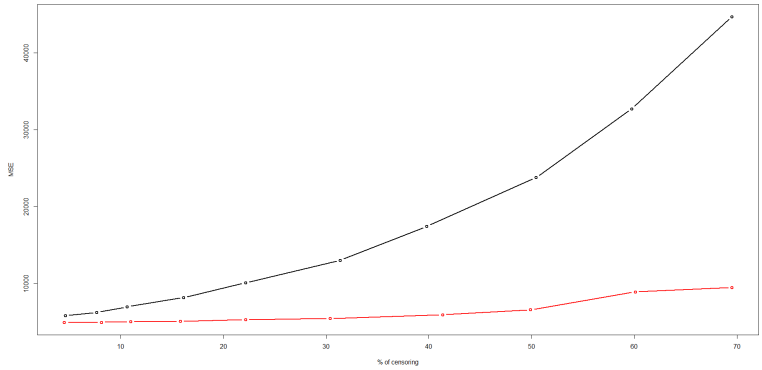

Figure 42: Mean squared error for different level of censoring in red with censoring taking into account and black without

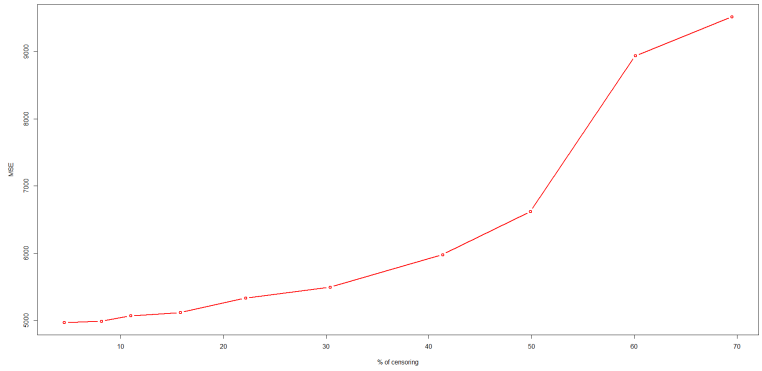

Figure 43: Mean squared error with censoring taking into account

The sum of square is significantly higher if the censoring is not taking into account. The sum of square start to show significant increase after 20% of censoring.

Taking into account censoring lead to better result. This especially interesting for the computation of the confidence interval of the mean. If censoring is taken into account, the value of the mean of the pain threshold increases with increasing censoring level. Conversely, if censoring is not taken into account, the value of the mean decreases as the censoring level increases.

9.2 The EM algorithm

The EM algorithm on simulated data can be computationally expensive, especially for a high level of censoring, but leads to a very accurate estimate. The computation is performed with the lmec package of R which is an optimal version of the EM algorithm in the study. The following table shows the result of the EM algorithm with different levels of censorship.

	True value	0%	5%	13%	23%	40%
Intercept	437.80051	508.9176	509.3067	509.8563	505.8806	496.4879
Time	11.59338	11.59338	11.59338	11.57990	11.57993	11.52423
Time	16.97063	16.97063	16.97063	16.95941	16.95719	16.83838
Time	14.74941	14.74941	14.74941	14.74169	14.73596	14.63462
Localisation	17.25923	17.25923	17.25861	17.25666	17.24507	17.13940
Localisation	56.39784	56.39784	56.39707	56.39400	56.39615	56.26496
Group	18.32266	-105.39162	-106.23791	-107.19470	-104.35884	-95.30308
Time×Group	-24.69131	-24.69131 7	-24.69065	-24.67800	-24.66906	-24.52309
Time×Group	-17.70270	-17.70270	-17.70270	-17.69141	-17.68604	-17.55684
Time×Group	-18.47402	-18.47402	-18.47402	-18.46385	-18.46899	-18.38024
sex	-144.97289	-159.3773	-159.5307	-159.7931	-157.5964	-154.6812
σ_ϵ	75.26	0.9995062	1.1475508	1.4338765	1.8039797	3.6852193
${\sigma_{\gamma 1}}^2$	26071	2310.13892	2240.41006	2172.17260	1885.21823	1680.00701
${\sigma_{\gamma 2}}^2$	1741	46.06586	45.40891	27.97303	29.44360	40.50042
${\sigma_{\gamma 3}}^2$	726	36.46706	36.71389	33.90755	23.91327	28.34155

The algorithm perfectly estimate the Time and Times× group parameters even with height level of censoring. Time variation is perfectly estimated. The impact of censoring is only visible on the other parameters. Without censoring taken into account the variation over time tend to fade as shown in figure 44.

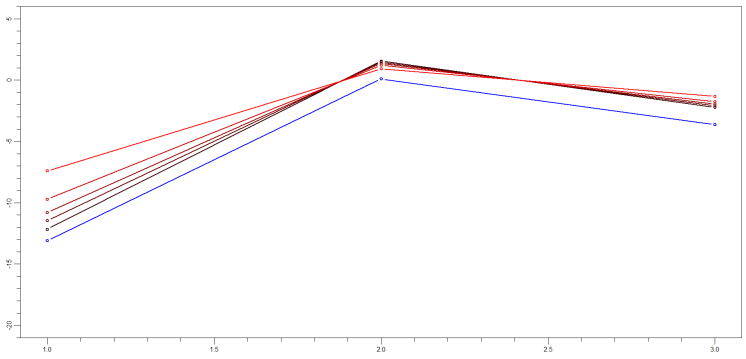

Figure 44: EM estimation of the Time+Timegroup parameter for different level of censoring without censoring taking into account

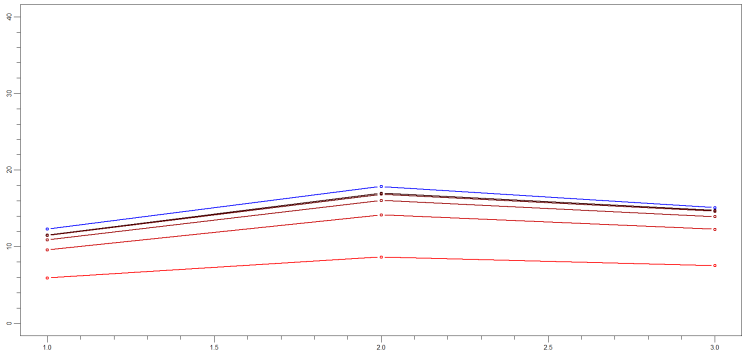

Figure 45: EM estimation of the Time parameter for different level of censoring without censoring taking into account

The EM algorithm taking into account censoring is more stable with increasing level of censoring.

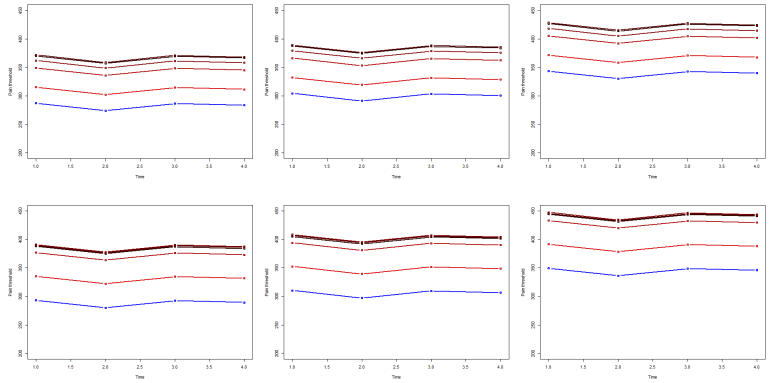

Figure 46: Mean of the regress pain threshold for the patient for different level of censoring with censoring taking into account

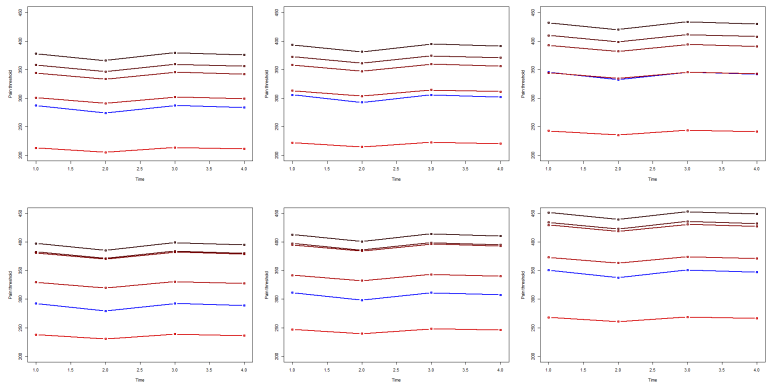

Figure 47: Mean of the regress pain threshold for the patient for different level of censoring without censoring taking into account

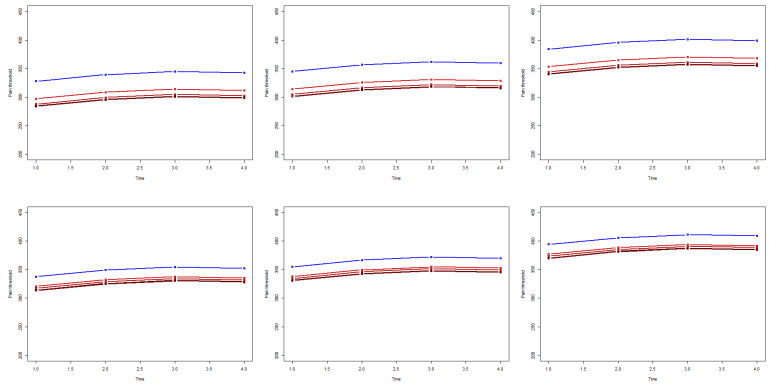

Figure 48: Mean of the regress pain threshold for the control for different level of censoring with censoring taking into account

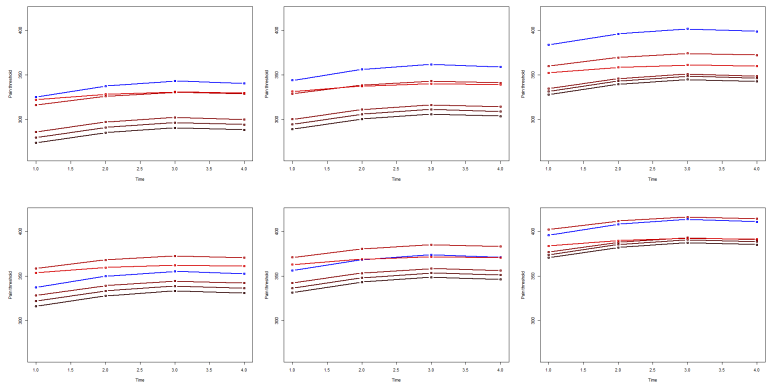

Figure 49: Mean of the regress pain threshold for the control for different level of censoring without censoring taking into account

10 Conclusion

Taking censoring into account provides a better estimate and a more stable result. Consideration of censoring gives a similar result to the initial study. A decrease in the pain threshold after exercise for the patient. This is due to the small amount of censored data (5 %) and a significant random error. The variation over time is also not significantly impacted by censoring. Censoring has a greater impact on intercept, group, location, and gender. This is particularly visible for censorship of 20% and above. The computational time needed to obtain a meaningful result with the EM algorithm with the study data increases significantly with the increase of the censoring level but gives very accurate results on simulated data even with a high censoring level.

References

[1] A Grimby-Ekman et al. "Pain intensity and pressure pain thresholds after a light dynamic physical load in patients with chronic neck-shoulder pain". In: *BMC musculoskeletal disorders* 21 (2020), pp. 1–10.

[2] James Tobin. "Estimation of relationships for limited dependent variables". In: *Econometrica: journal of the Econometric Society* (1958), pp. 24–36.

[3] Giuseppe Bruno et al. "Limited dependent panel data models: a comparative analysis of classical and Bayesian inference among econometric packages". In: *Computing in Economics and Finance*. Vol. 41. 2004.

[4] Seung-Chun Lee and Byongsu Choi. "Bayesian inference for censored panel regression model". In: *Communications for Statistical Applications and Methods* 21.2 (2014), pp. 193–200.

[5] CS Wang, JJ Rutledge, and D Gianola. "Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs". In: *Genetics Selection Evolution* 26.2 (1994), pp. 91–115.

[6] Nan M Laird and James H Ware. "Random-effects models for longitudinal data". In: *Biometrics* (1982), pp. 963–974.

[7] James P Hughes. "Mixed effects models with censored data with application to HIV RNA levels". In: *Biometrics* 55.2 (1999), pp. 625–629.

[8] Florin Vaida, Anthony P Fitzgerald, and Victor DeGruttola. "Efficient hybrid EM for linear and nonlinear mixed effects models with censored response". In: *Computational statistics & data analysis* 51.12 (2007), pp. 5718–5730.

[9] Amemiya Takeshi. *Advanced econometrics*. Harvard university press, 1985.

[10] William H Greene. "Econometric analysis 6th edition". In: *International edition, New Jersey: Prentice Hall* (2000).

[11] David Harville. "Extension of the Gauss-Markov theorem to include the estimation of random effects". In: *The Annals of Statistics* (1976), pp. 384–395.

[12] John S Butler and Robert Moffitt. "A computationally efficient quadrature procedure for the one-factor multinomial probit model". In: *Econometrica: Journal of the Econometric Society* (1982), pp. 761–764.