Study 2. Multicentric Randomized Clinical Trial - Fixed vs. Mixed Effect Model

Ting Lu

2023-10-24

Introduction

In a double-blinded multicentric Randomized Clinical Trial (RCT with 36 centers), a total of 298 sportsmen and elderly people were treated for **toenail dermatophyte onychomycosis** with either of two oral medications: Itraconazol 250 mg daily (treat=0) or Lamisil 250mg daily (treat=1). The patients received treatment for 12 weeks and were evaluated at 0, 1, 2, 3, 6, 9 and 12 months. As response, the unaffected nail length for the big toenail was measured.

The study wants to determine whether the treatment works on average and if there is a difference among the treatments.

Install Package

```
# library(tidyverse): data manipulation & visualization
# library(MASS): statistical methods & regression analysis
# library(MCMCpack): MCMC simulations &  Bayesian statistical modeling and parameter estimation
```

`Tidyverse` includes multiple packages: `ggplot2`, `dplyr`, `tidyr`, `readr`, `purrr`, `tibble`, `stringr`

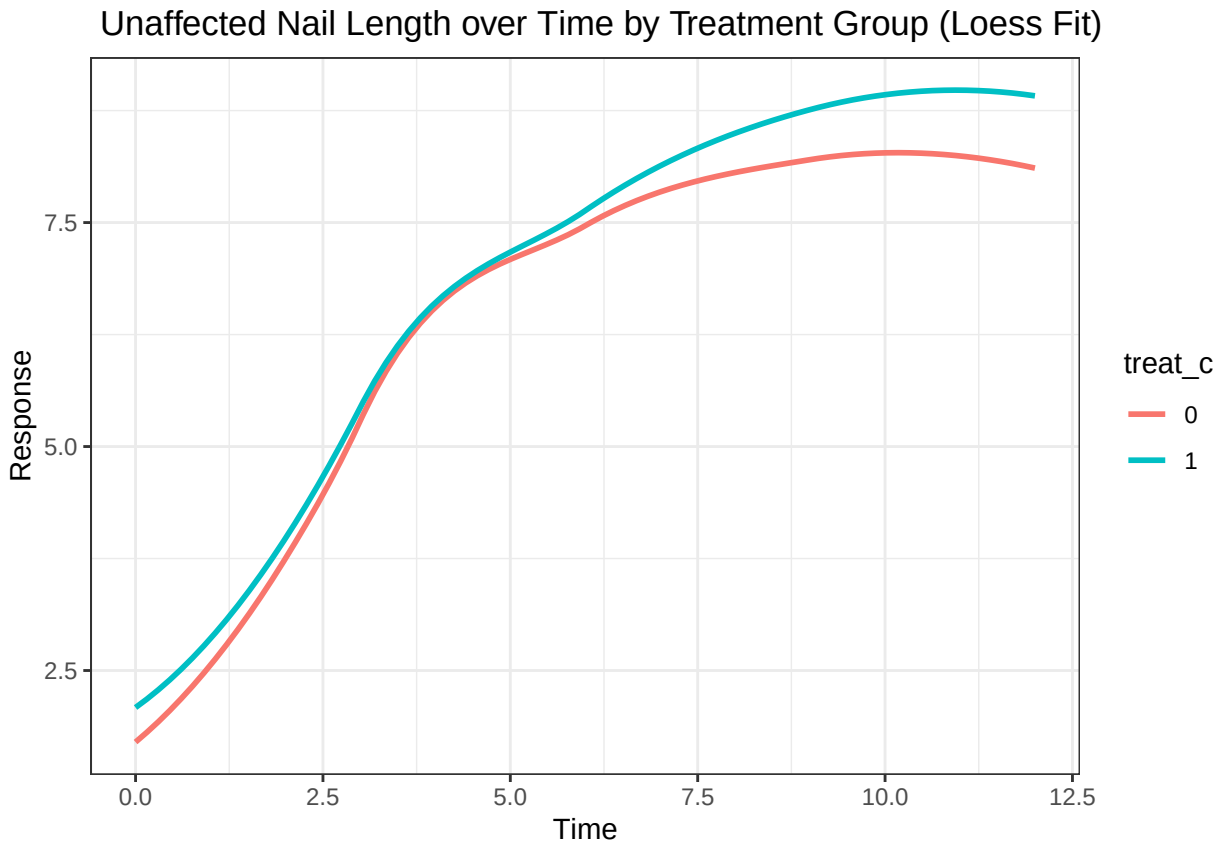
Loading data

```
toe = read.table("/Users/tinglu/Bayesian/Study2-Multicenter-RCT/toenail.txt", header=T)
str(toe)
```

```
## 'data.frame':    1854 obs. of  5 variables:
##  $ obs     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ treat   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ id      : int  2 2 2 2 2 2 3 3 3 3 ...
##  $ time    : int  0 1 2 3 6 9 0 1 2 3 ...
##  $ response: num  4 6 7 9 13 0 0 1 2 3 ...
```

EDA (Response vs. Time)

```
toe$treat_c=as.character(toe$treat)
ggplot(toe, aes(x = time, y = response, color = treat_c)) +
  geom_smooth(method = "loess", formula= y ~ x, se = FALSE) +
  labs(x = "Time", y = "Response") +
  ggtitle(" Unaffected Nail Length over Time by Treatment Group (Loess Fit)") +
  theme_bw()
```



EDA (OLS regression for two treatment groups)

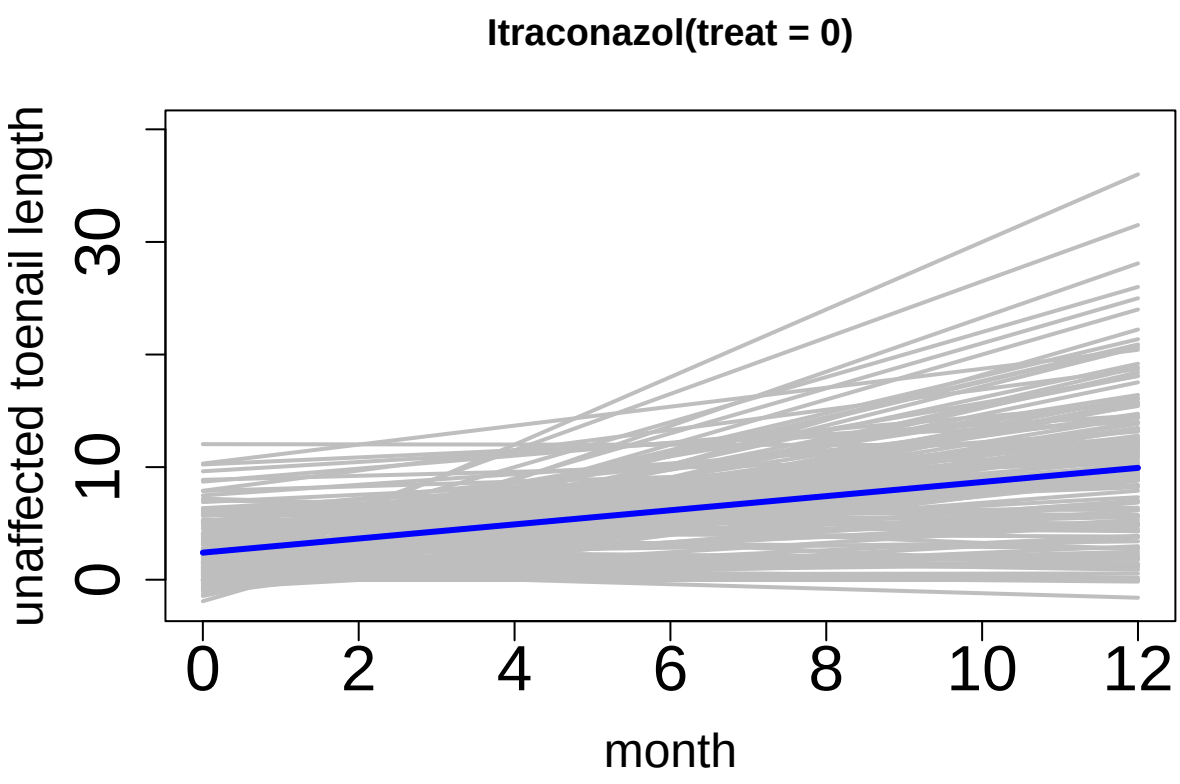
Because each patients has his/her own physiological and biological characteristics and might respond to treatment differently, conduct an exploratory analysis by fitting the OLS regression model separately for each patient i within two treatment groups:

$$y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma_i^2)$$

For Itraconazol group (treat = 0)

```
# For Itraconazol group (treat = 0)
toe0 = toe[toe$treat == 0,]
```

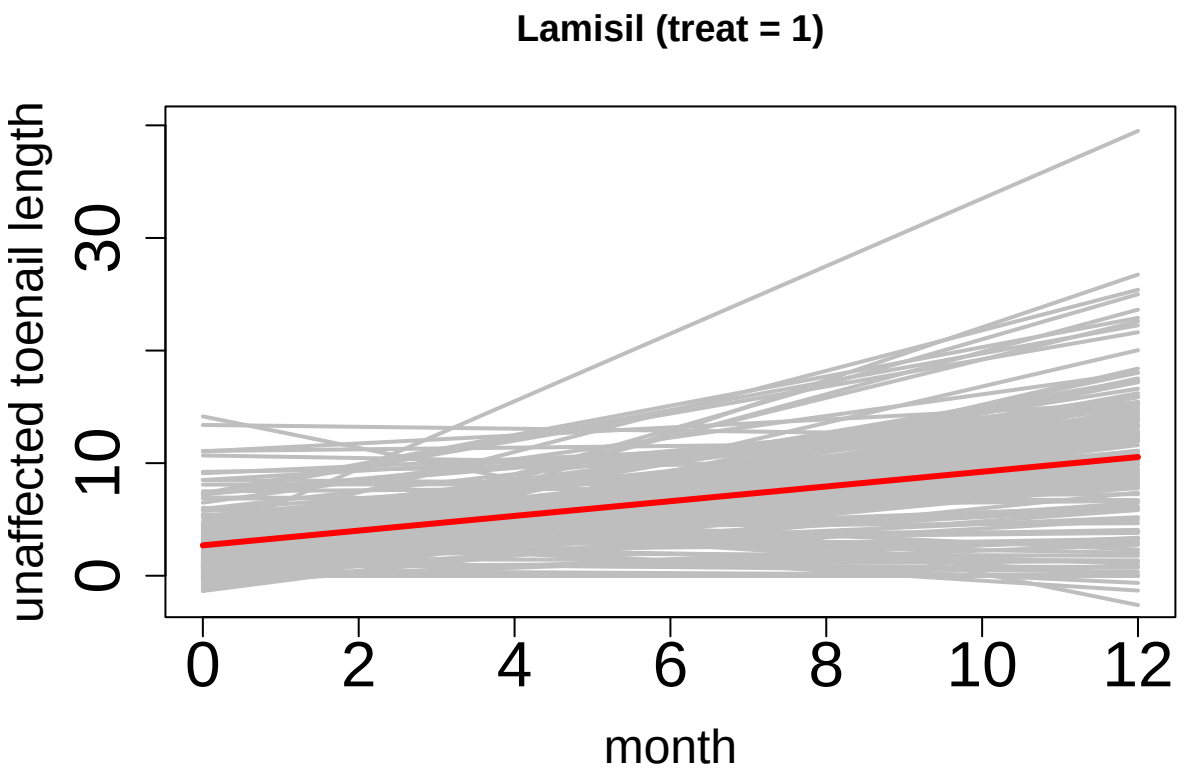
```r
pat <- toe0$id
m <- length(unique(pat))
y <- toe0$response
n <- length(y)
d <- toe0$time
p <- 2
X <- matrix(cbind(rep(1,n),d),nrow=n,ncol=p)

## determining the number of observations for each group(each patients)
n.j <- as.numeric(table(pat))
## OLS estimates of regression coefficients
beta.ols <- matrix(0,nrow=m,ncol=p)
var.pat <- rep(0,m)
sigma2 <- NULL
for(j in 1:m){
    y.j <- y[which(pat==unique(pat)[j])]
    d.j <- d[which(pat==unique(pat)[j])]
    reg.j <- lm(y.j~d.j)
    beta.ols[j,] <- as.numeric(reg.j$coeff)
    var.pat[j] <- var(y.j)
    sigma2[j] <- deviance(reg.j)/reg.j$df.residual
}
d_index = seq(0,12, length.out = 12)
plot(d_index,beta.ols[1,1]+beta.ols[1,2]*d_index,xlab="month",ylab="unaffected toenail length",col="grey

for(j in 2:m){
 lines(d_index,beta.ols[j,1]+beta.ols[j,2]*d_index,col="grey",lwd=2)
}
lines(d_index,mean(beta.ols[,1],na.rm = T)+mean(beta.ols[,2],na.rm = T)*d_index,col="blue",lwd=3)
```



```
beta.ols0 <- beta.ols
```

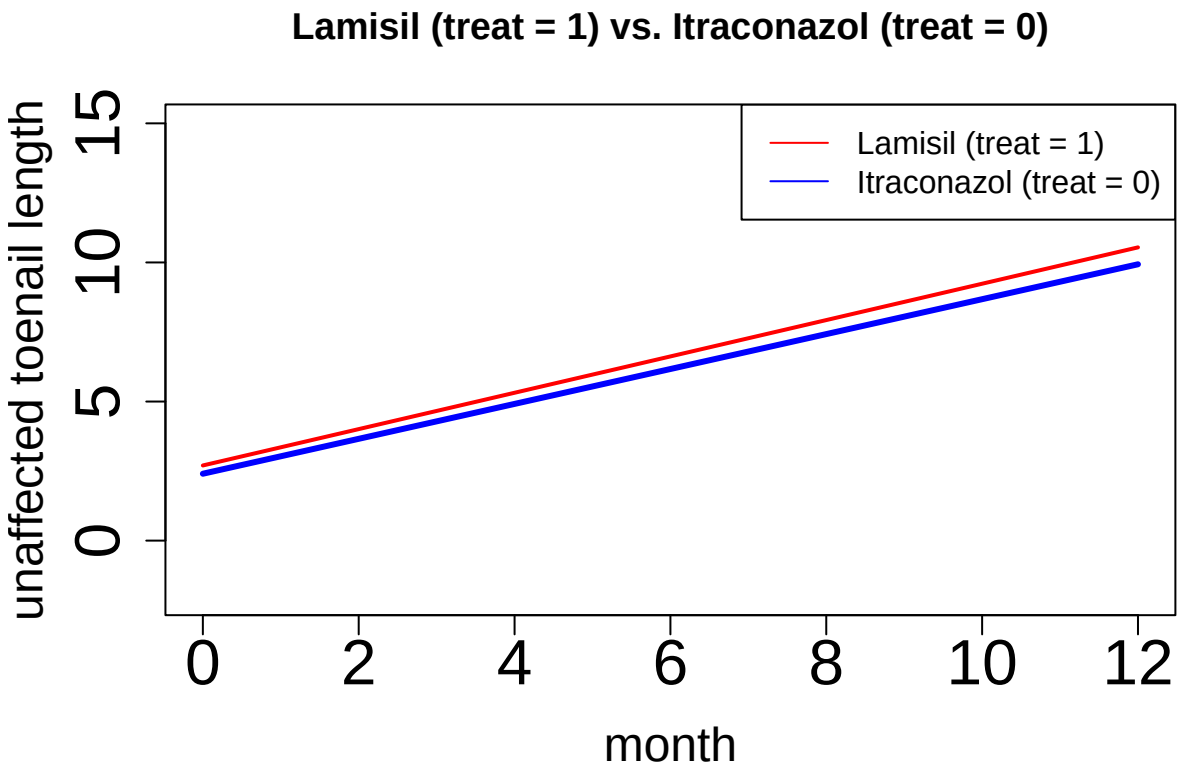
For Lamisil group (treat = 1)

```
# For Lamisil group (treat = 1)
toe1 = toe[toe$treat == 1,]
pat <- toe1$id
m <- length(unique(pat))
y <- toe1$response
n <- length(y)
d <- toe1$time
p <- 2
X <- matrix(cbind(rep(1,n),d),nrow=n,ncol=p)

## determining the number of observations for each group(each patients)
n.j <- as.numeric(table(pat))
## OLS estimates of regression coefficients
beta.ols <- matrix(0,nrow=m,ncol=p)
var.pat <- rep(0,m)
sigma2 <- NULL
for(j in 1:m){
    y.j <- y[which(pat==unique(pat)[j])]
    d.j <- d[which(pat==unique(pat)[j])]
    reg.j <- lm(y.j~d.j)
    beta.ols[j,] <- as.numeric(reg.j$coeff)
    var.pat[j] <- var(y.j)
    sigma2[j] <- deviance(reg.j)/reg.j$df.residual
}
d_index = seq(0,12, length.out = 12)
plot(d_index,beta.ols[1,1]+beta.ols[1,2]*d_index,xlab="month",ylab="unaffected toenail length",col="grey

for(j in 2:m){
 lines(d_index,beta.ols[j,1]+beta.ols[j,2]*d_index,col="grey",lwd=2)
}
lines(d_index,mean(beta.ols[,1],na.rm = T)+mean(beta.ols[,2],na.rm = T)*d_index,col="red",lwd=3)
```



Lamisil (treat = 1) vs. Itraconazol (treat = 0)

```
plot(d_index,mean(beta.ols[,1],na.rm = T)+mean(beta.ols[,2],na.rm = T)*d_index,xlab="month",ylab="unaff
lines(d_index,mean(beta.ols0[,1],na.rm = T)+mean(beta.ols0[,2],na.rm = T)*d_index,col="blue",lwd=3)
legend("topright", legend = c("Lamisil (treat = 1)","Itraconazol (treat = 0)"), col = c("red","blue"),l
```



In either the Itraconazol(treat $= 0$) or the Lamisil(treat $= 1$) group, the average evolution of unaffected toenail length over time grows. Regression line of the Lamisil group is slightly higher than that of Itraconazol group.

In either the Itraconazol(treat $= 0$) or the Lamisil(treat $= 1$) group, **each individual has different regression line**, indicating **much variability**. For most of individuals, evolution of toenail length over time grows. However, for some other individuals, evolution of toenail length over time decreases, so it is reasonable to compare two models:

- **linear fixed effect model**
- **linear mixed model (with random slope and intercept)** to fit each patients' own physiological and biological characteristics.

1. Linear fixed-effect regression model:

It is necessary to take both the time factor and the treatment drug factor into consideration. Firstly, establish a linear **fixed-effect** regression model.

$$y_{ij} = \beta_0 + \beta_1 * t_{ij} + \beta_2 * (t_{ij} * treat_i) + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

-y_{ij}: outcome of patient i at time $t_{ij} = 0, 1, 2, 3, 6, 9, 12$ months under treatment $treat_i$.
-No main effect for treatment was included in the model since at baseline the effect of treatment must be zero given the randomized character of the study. This implies that at the beginning of the study, there should be no difference in toenail length between the treatment groups.

Transform the equation, it is a fixed effect that impact the slope:

$$y_{ij} = \beta_0 + (\beta_1 + \beta_2 * treat_i)t_{ij} + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

Prior setting & Initial value

Fit the above fixed effect model using the following priors based on previous studies knowledge and observed data:

```
knitr::include_graphics("/Users/tinglu/Bayesian/Study2-Multicenter-RCT/toe.png",error = F)
```

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 2.5 \\ 0.6 \\ 0 \end{pmatrix}, 100 \times \mathbf{I}_3 \right)$$

$$\sigma^2 \sim \text{Inverse Gamma}(1, 3.7)$$

Gibbs sampling (MCMC sampling) for fixed-effect regression model

Run Gibbs sampling, using observed data to update the prior information and derive posterior means and 95% credible intervals for the regression coefficients and for s2 and determine whether

- (i) the unaffected toe nail grows significantly over time;
- (ii) there is a difference between the two treatments over time.(Itraconazol 250 mg daily (treat=0) vs. Lamisil 250mg daily (treat=1).)

```
# Sampling parameter
S <- 10000
burn_in <- 1000
X <- cbind(1, toe$time, toe$time * toe$treat)
y <- toe$response
n <- nrow(X)
p <- ncol(X)
# prior
beta0 <- c(2.5, 0.6, 0)
sigma0 <- diag(rep(100, 3))
nu0 <- 2
s20 <- 3.7
# initial values
beta <- beta0
s2 <- s20

Fix.beta <- matrix(nrow = S, ncol = length(beta0))
Fix.sigma <- numeric(S)
set.seed(0)
```

```
# Gibbs sampling
for (s in 1:S) {
  V <- solve(solve(sigma0) + (t(X) %*% X)/s2)
  m <- V %*% (solve(sigma0) %*% beta0 + (t(X) %*% y)/s2)
  beta <- mvrnorm(1, m, V)
  ssr <- (t(y) %*% y) - (2 * t(beta) %*% t(X) %*% y) + (t(beta) %*%
                                                        t(X) %*% X %*% beta)
  s2 <- 1/rgamma(1, (nu0 + n)/2, (nu0 * s20 + ssr)/2)

  Fix.beta[s, ] <- beta
  Fix.sigma[s] <- s2
}
```

Posterior inference for linear fixed-effect model

```
post.mean.fix <- apply(cbind(Fix.beta[burn_in:S, ], Fix.sigma[burn_in:S]), 2, mean)
post.quantile.fix <- apply(cbind(Fix.beta[burn_in:S, ], Fix.sigma[burn_in:S]), 2, quantile,
c(0.025, 0.975))
fix <- rbind(post.mean.fix, post.quantile.fix)
colnames(fix) <- c("beta.(Intercept)", "beta.time", "beta.time.treat", "sigma2")
fix
```

```
##               beta.(Intercept) beta.time beta.time.treat   sigma2
## post.mean.fix         2.664289 0.5609139      0.06346246 13.55644
## 2.5%                  2.423144 0.5096115      0.00667572 12.70317
## 97.5%                 2.909417 0.6125981      0.12015786 14.45544
```

linear fixed-effect model

$$y_{ij} = \beta_0 + (\beta_1 + \beta_2 * treat_i)t_{ij} + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

Both 95% credible intervals of time and the interaction term of treat and time don't include 0, indicating unaffected toe nail grows significantly over time and there is **significant difference** between the two treatments over time.

2. Hierachical Linear Mixed Model

$$y_{ij} = \beta_{0i} + \beta_{1i} * t_{ij} + \beta_2 * (t_{ij} * treat_i) + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

Prior setting for Hierachical Linear Mixed Model

$$\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} | \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \Sigma \overset{iid}{\sim} N_2\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \Sigma\right) \qquad i = 1, \ldots, 298$$

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 2.5 \\ 0.6 \end{pmatrix}, 100 \times \mathbf{I}_2\right)$$

$$\Sigma \sim \text{Inverse Wishart}(4, 0.1 \times \mathrm{I}_2)$$

$$\beta_2 \sim N(0, 100)$$

$$\sigma^2 \sim \text{Inverse Gamma}(1, 3.7)$$

```
# Normal, for beta
mu0 <- c(2.5, 0.6, 0)
Sigma0 <- diag(rep(100, 3))
init.mu <- mu0
init.Sigma <- Sigma0
# Inverse Gamma, for residuals
nu0 <- 1
delta0 <- 3.7
# Inverse Wishart, for random effect
r <- 4
R <- 0.1* diag(rep(1, 2))/r
toe$time.treat <- toe$time*toe$treat
```

Gibbs sampling (MCMC sampling)

```
model <- MCMChregress(fixed = response ~ time + time.treat,
                      random = ~time,
                      group = "id",
                      data = toe,
                      burnin = 1000, mcmc = 10000,
                      thin = 1, verbose = 1, seed = 0,
                      mubeta = init.mu, Vbeta = init.Sigma,
                      r = r, R = R,
                      nu = nu0, delta0 = 3.7)
```

Using "MCMCpack" package

```
##
## Running the Gibbs sampler. It may be long, keep cool :)
##
## **********:10.0%
## **********:20.0%
## **********:30.0%
## **********:40.0%
## **********:50.0%
## **********:60.0%
## **********:70.0%
## **********:80.0%
## **********:90.0%
## **********:100.0%
```

MCMC sampling diagnostic (trace plot, ACF plot, effective sampling size)

```
dim(model$mcmc) # contains all posterior samples
```

To make sure Markov chain is converged.

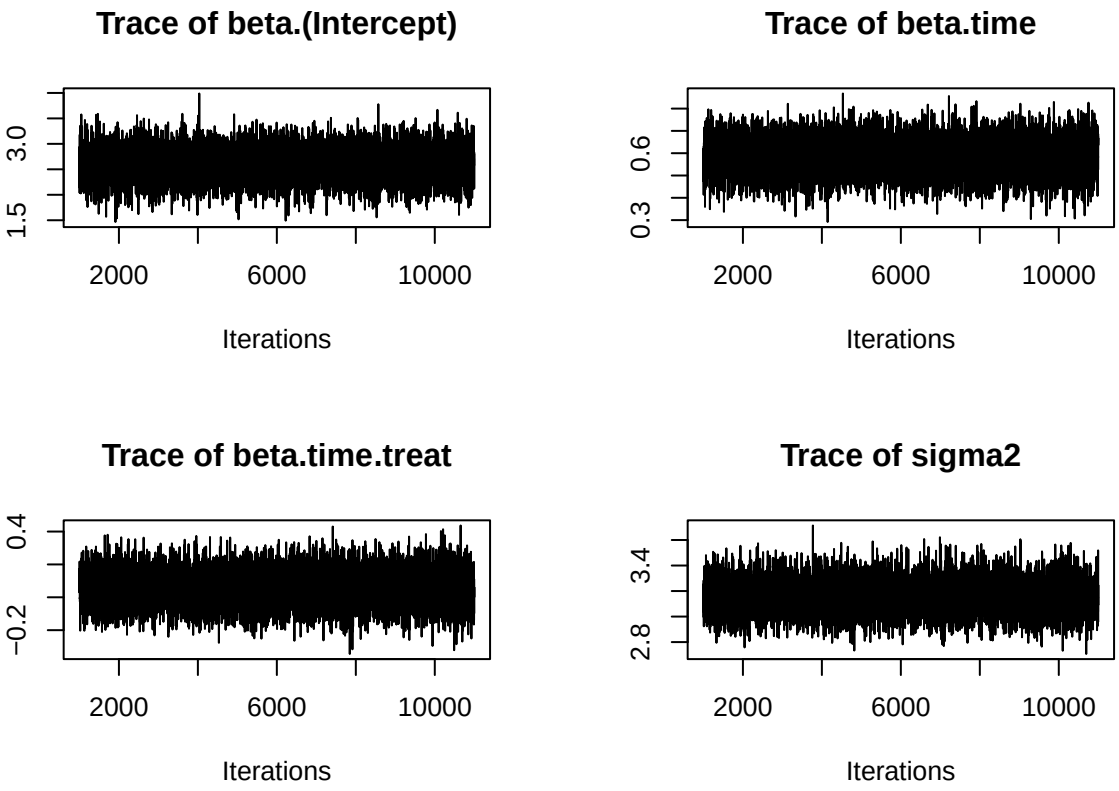
```
## [1] 10000   605
```

```
sum(sapply(1:605, function(i) effectiveSize(model$mcmc[,i]))<=1000)
```

```
## [1] 0
```

Effect sample size (ESS) of all parameters is large than 1000, which means enough samples provide independent information, increasing the reliability of estimates.

```
par(mfrow = c(2,2))
traceplot(model$mcmc[, c(1:3, 604)], smooth = TRUE)
```



Assess convergence for $\beta_0, \beta_1, \beta_2, \sigma^2$, and Trace plot show the Markov chain is converged.

Posterior inference (Hierachical Linear Mixed Model)

```
post.mean.mix <-apply(model$mcmc[, c(1:3, 604)][1000:10000, ], 2, mean)
post.quantile<-apply(model$mcmc[, c(1:3, 604)][1000:10000, ], 2, quantile, c(0.025, 0.975))
mixed <- rbind(post.mean.mix,post.quantile)
mixed
```

```
##               beta.(Intercept) beta.time beta.time.treat   sigma2
## post.mean.mix         2.596517 0.5823375      0.05661579 3.155056
## 2.5%                  2.002294 0.4319638     -0.14672889 2.915619
## 97.5%                 3.188220 0.7329455      0.25761339 3.411383
```

95% credible interval of β_1 (time) don't include 0, which means average unaffected toenail length grows over time.

95% credible interval of β_2 (time*treat) include 0, which means there is no significant difference between these two treatment groups.

Fixed effect model vs. Mixed effect model

```
fix
```

```
##               beta.(Intercept) beta.time beta.time.treat   sigma2
```

```
## post.mean.fix         2.664289 0.5609139      0.06346246 13.55644
## 2.5%                  2.423144 0.5096115      0.00667572 12.70317
## 97.5%                 2.909417 0.6125981      0.12015786 14.45544
```

```
mixed
```

```
##               beta.(Intercept) beta.time beta.time.treat   sigma2
## post.mean.mix         2.596517 0.5823375      0.05661579 3.155056
## 2.5%                  2.002294 0.4319638     -0.14672889 2.915619
## 97.5%                 3.188220 0.7329455      0.25761339 3.411383
```

Comparing the Fixed Effect Model and Mixed Effect Model, σ^2_{fix} is significantly greater than σ^2_{mixed}, so there is variability among patients in the way their unaffected toenail(response) grows over time. This implies that in the study, there is more variability in unaffected toenail length (the response variable) among patients over time in the Fixed Effect Model, while there is less variability among patients in the Mixed Effect Model.

1. **Fixed Effect Model**:
    - In the Fixed Effect Model, σ^2_{fix} represents the overall variance in toenail length change among all patients. This model assumes that the differences between all patients are explained by random error terms e_{ij} and does not consider that the differences between patients are caused by individual variability.
    - This means that the Fixed Effect Model does not account for the random variability between patients and treats all patients as coming from the same overall distribution. Therefore, the model may overestimate the overall variance in unaffected toenail length.
2. **Mixed Effect Model**:
    - In the Mixed Effect Model, σ^2_{mixed} takes into account the differences between patients. This model divides the variation in unaffected toenail length into two parts: fixed effects and random effects.
    - Random effects capture the variability between different patients. In the Mixed Effect Model, a portion of the variance is attributed to between-patient differences (random effects), while another portion of the variance is caused by random error terms.
    - This allows the Mixed Effect Model to more accurately describe the differences between individual patients, resulting in a smaller variance in σ^2_{mixed}.

In summary, the significant difference is because the Mixed Effect Model more reasonably considers individual differences between patients, while the Fixed Effect Model assumes that all patients come from the same overall distribution and fails to capture patient-to-patient variability. The Mixed Effect Model is more suitable for describing data that includes random effects.

Prediction (Hierachical Linear Mixed Model)

For two new patients (Itraconzol vs.Lamisil), predict how long their unaffected toe nail will be at 0, 1, 2, 3, 6, 9 and 12 months since enrollment in the study and provide 95% confidence bands:

```
beta <- model$mcmc[1000:10000, 1:3]
ti <- c(0,1,2,3,6,9,12)
X0 <- cbind(1,ti,ti*0)
X1 <- cbind(1,ti,ti*1)
pred0 <- X0 %*% t(beta)
pred1 <- X1 %*% t(beta)
```

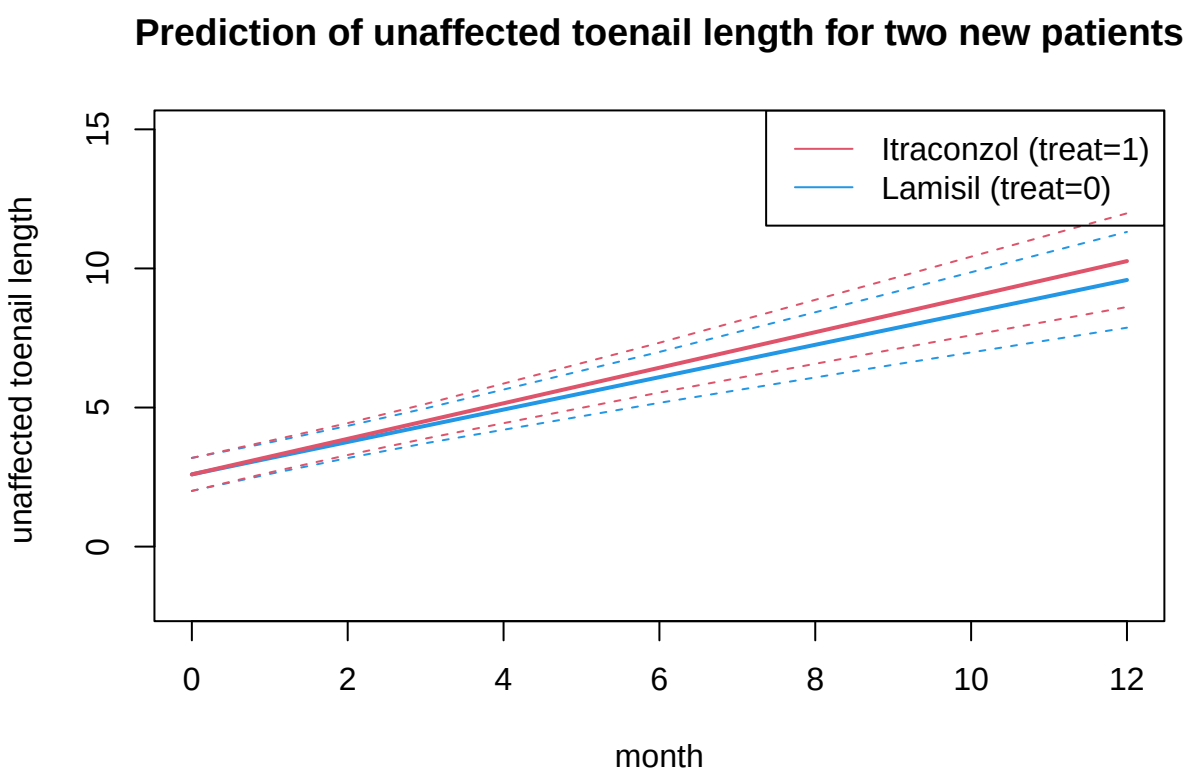
```r
pred0.mean <- apply(pred0, 1, mean)
pred0.quantile <- apply(pred0, 1, quantile, c(0.025, 0.975))
pred1.mean <- apply(pred1, 1, mean)
pred1.quantile <- apply(pred1, 1, quantile, c(0.025, 0.975))
df0 = data.frame(time = ti, Q2.5 = round(pred0.quantile[1, ],2),mean = round(pred0.mean,2), Q97.5 = rou
df1 = data.frame(time = ti,Q2.5 = round(pred1.quantile[1, ],2),mean = round(pred1.mean,2), Q97.5 = roun
# prediction for treat 0 - Itraconzol
df0
```

```
##   time Q2.5 mean Q97.5
## 1    0 2.00 2.60  3.19
## 2    1 2.61 3.18  3.75
## 3    2 3.18 3.76  4.34
## 4    3 3.72 4.34  4.97
## 5    6 5.17 6.09  7.00
## 6    9 6.54 7.84  9.14
## 7   12 7.87 9.58 11.31
```

```r
# prediction for treat 1 - Lamisil
df1
```

```
##   time Q2.5  mean Q97.5
## 1    0 2.00  2.60  3.19
## 2    1 2.67  3.24  3.80
## 3    2 3.29  3.87  4.44
## 4    3 3.89  4.51  5.13
## 5    6 5.54  6.43  7.33
## 6    9 7.09  8.35  9.64
## 7   12 8.61 10.26 11.98
```

```r
df0 = data.frame(time = ti,low = pred0.quantile[1, ],mean = pred0.mean, high = pred0.quantile[2, ])
df1 = data.frame(time = ti,low = pred1.quantile[1, ],mean = pred1.mean, high = pred1.quantile[2, ])

plot(df0$time,df0$mean,type= 'l',xlab = 'month', ylab="unaffected toenail length ",ylim=c(-2,15),col =
lines(df0$time,df0$low,type = 'l', lty = 2,col = '4')
lines(df0$time,df0$high,type = 'l',lty = 2,col = '4')

lines(df1$time,df1$mean,type = 'l', lty = 1,col = '2',lwd=2)
lines(df1$time,df1$high,type = 'l',lty = 2,col = '2')
lines(df1$time,df1$low,type = 'l', lty = 2,col = '2')

legend("topright", legend = c("Itraconzol (treat=1)","Lamisil (treat=0)"), col = c(2,4),lty=1,cex=1)
```


Prediction of unaffected toenail length for two new patients
unaffected toenail length
0
5
10
15
0
2
4
6
8
10
12
month
Itraconzol (treat=1)
Lamisil (treat=0)