

Final Homework Mixed Effects Models

Jordan Bonil

April 2022

Part1

0.1

We import the data with the function `read.table`. We also remove the rows which contain missing values with the function `na.omit`.

We visualize the QQ Plot of the raw data in column CD4. Then we transform the column to see whether the QQ plot shows improvement :

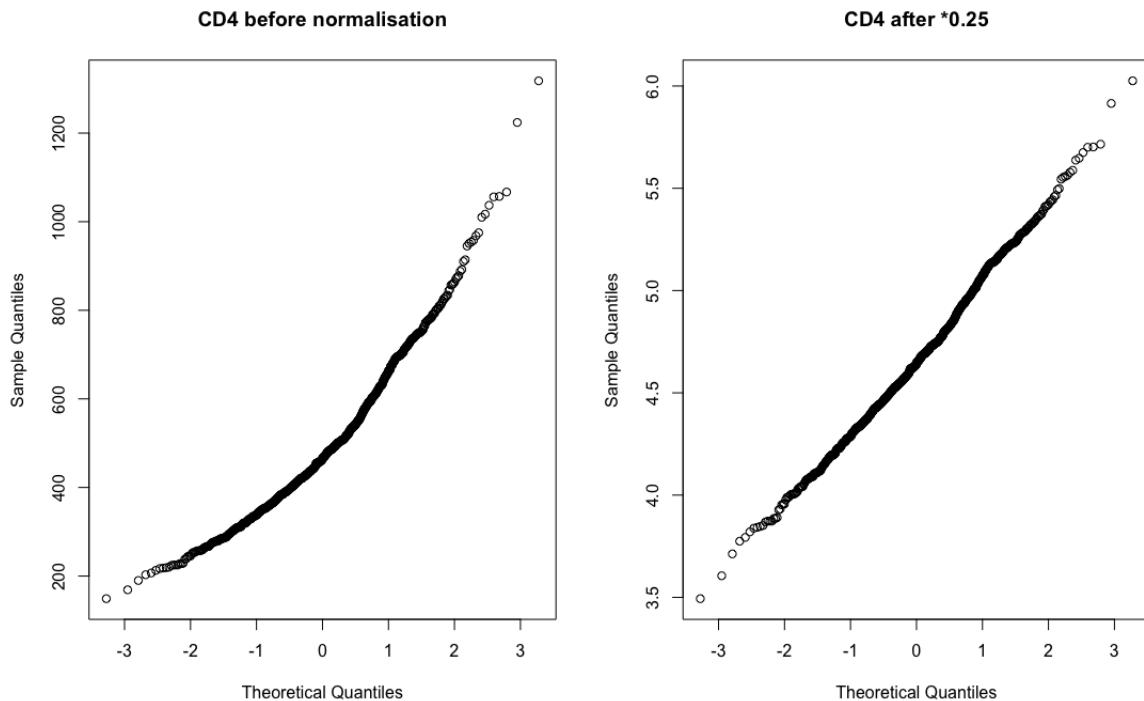


Figure 1: QQ Plots before and after a $^{.0.25}$ normalisation

The data in the VL column acts different. We apply a \log_{10} transformation, and notice a that the datapoints below the third quartile mess up the QQ Plot :

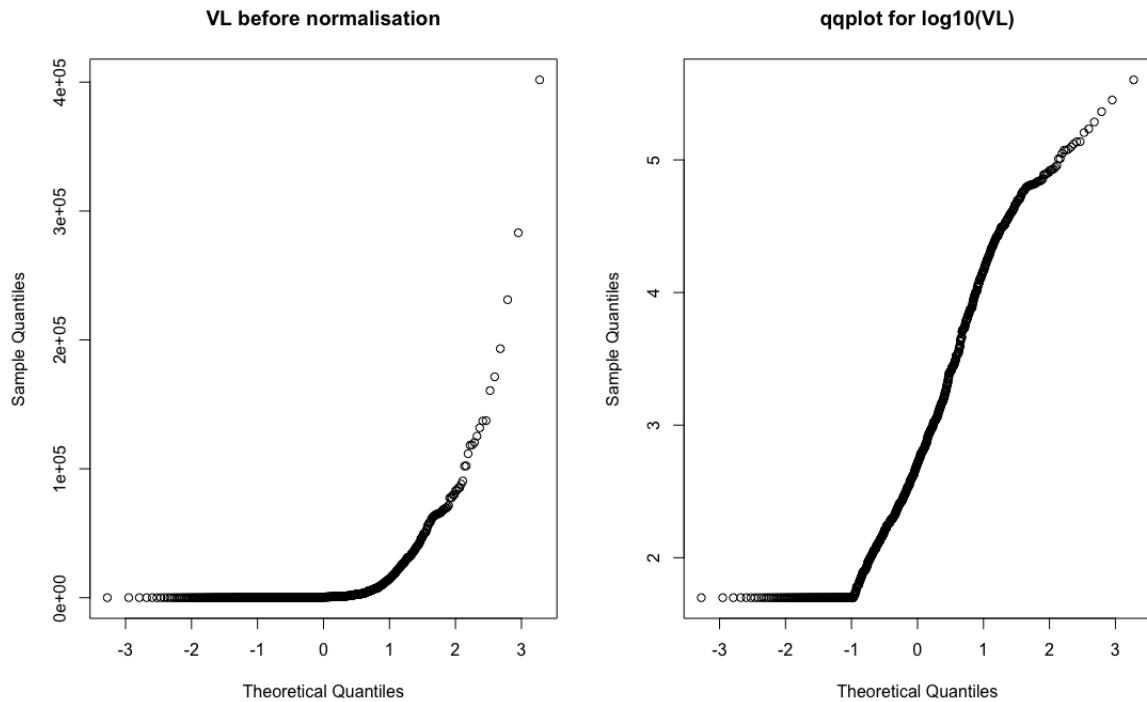


Figure 2: QQ Plots before and after a \log_{10} normalisation

We will then apply a second transformation, by thresholding the datapoints above the 3rd quantile. The resulting QQ Plot shows the improvement over original data :

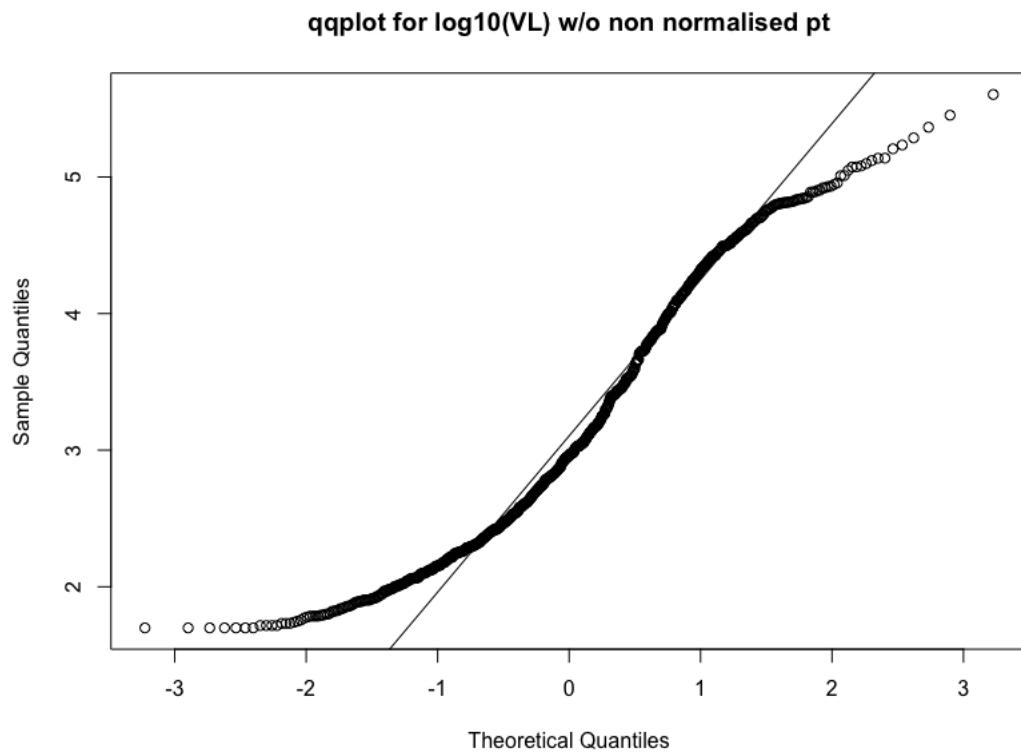


Figure 3: Normalisation is improved while still imperfect

Moving on, in order to create a new variable as explained in the instructions, we use the command
`data$dd4Tddl = as.numeric(data$RAN_GRP != 1)`

0.2

```
> summary(data)
      NUM_PAT          TIME           VIS            TD
Min.   : 15.0   Min.   : 0.00   Min.   :0.000   Min.   : 0.00
1st Qu.:239.0  1st Qu.: 28.00  1st Qu.:1.000  1st Qu.: 4.00
Median :492.0   Median : 84.00  Median :3.000  Median :12.00
Mean   :489.5   Mean   : 81.53  Mean   :2.783  Mean   :11.13
3rd Qu.:734.0  3rd Qu.:138.00 3rd Qu.:5.000  3rd Qu.:20.00
Max.   :997.0   Max.   :256.00  Max.   :6.000  Max.   :24.00
      RAN_GRP          CD4          VL            dd4Tddl
Min.   :1.000   Min.   :3.494   Min.   :1.699   Min.   :0.0000
1st Qu.:1.000  1st Qu.:4.392   1st Qu.:2.333  1st Qu.:0.0000
Median :2.000   Median :4.631   Median :2.961   Median :1.0000
Mean   :1.914   Mean   :4.658   Mean   :3.135   Mean   :0.6245
3rd Qu.:3.000  3rd Qu.:4.918   3rd Qu.:3.876  3rd Qu.:1.0000
Max.   :3.000   Max.   :6.025   Max.   :5.604   Max.   :1.0000
```

The data contains TIME and TD columns, which correspond respectively to number of days and number of weeks since the treatment began. The columns NUM_PAT, RAN_GRP, dd4tddl and VIS are all categorical, so that usual statistics (mean, standard dev) are meaningless. Instead, we may list the following points of interests :

- there are 149 patients dispatched uniformly in 3 groups
- 3 different treatments
- 256 days is the max elapsed time of observation

To obtain the average trajectories of VL and CD4, we follow this procedure : on each given week, if more than one concentration point is available, we average out the concentrations. If not, we take the one concentration.

Average trajectory for Viral Load wrt to Time

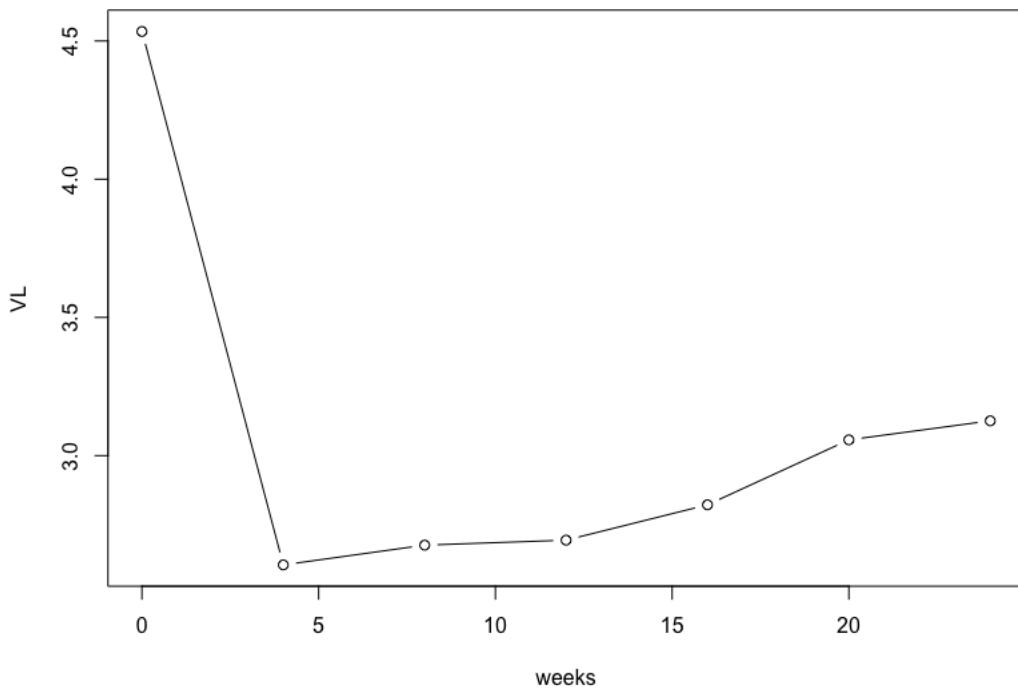


Figure 4: We notice that the Viral Load drops significantly very early on, and then slowly trails up by a few decimal points

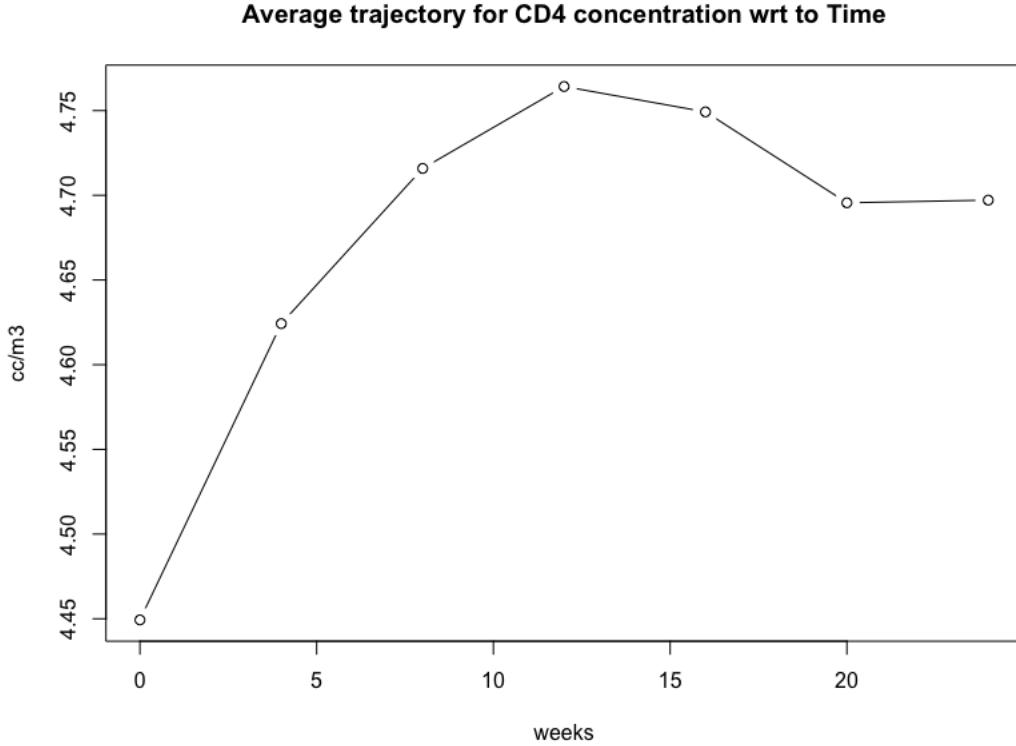


Figure 5: This graph shows how CD4s increase rapidly at first, with a max value reached a few weeks in, then plateaus

0.3

Let us now focus on RAN_GRP 1 and 3 with the command `sub_grp = subset(data, data$RAN_GRP != 2)`. We import the Mixed Effects library `lme4` to run a few tests on explaining CD4 count as a linear function of TD and TD^2 . This is done with the following commands :

```
library(lme4)

# trying to find the best model :
m1 <- lm(sub_grp$CD4 ~ sub_grp$TD2 , data=sub_grp)
m2 <- lm(sub_grp$CD4 ~ sub_grp$TD + sub_grp$TD2 , data=sub_grp)
m3 <- lm(sub_grp$CD4 ~ 1 + sub_grp$TD2:sub_grp$TD , data=sub_grp)
m4 <- lm(sub_grp$CD4 ~ sub_grp$TD + sub_grp$TD2:sub_grp$TD , data=sub_grp)
m5 <- lmer(sub_grp$CD4 ~ sub_grp$TD2 + (1|sub_grp$NUM_PAT) , data=sub_grp)
m6 <- lmer(sub_grp$CD4 ~ sub_grp$TD + sub_grp$TD2 + (1|sub_grp$NUM_PAT) , data=sub_grp)
m7 <- lmer(sub_grp$CD4 ~ 1 + sub_grp$TD2:sub_grp$TD + (1|sub_grp$NUM_PAT) , data=sub_grp)
m8 <- lmer(sub_grp$CD4 ~ sub_grp$TD + sub_grp$TD2:sub_grp$TD + (1|sub_grp$NUM_PAT) , data=sub_grp)
m9 <- lmer(sub_grp$CD4 ~ sub_grp$TD2 + (-1+sub_grp$TD2|sub_grp$NUM_PAT) , data=sub_grp)
m10 <- lmer(sub_grp$CD4 ~ sub_grp$TD + sub_grp$TD2 + (-1+sub_grp$TD2|sub_grp$NUM_PAT) , data=sub_grp)
m11 <- lmer(sub_grp$CD4 ~ 1 + sub_grp$TD2:sub_grp$TD + (-1+sub_grp$TD2|sub_grp$NUM_PAT) , data=sub_grp)
m12 <- lmer(sub_grp$CD4 ~ sub_grp$TD + sub_grp$TD2:sub_grp$TD + (-1+sub_grp$TD2|sub_grp$NUM_PAT) , data=sub_grp)
m13 <- lmer(sub_grp$CD4 ~ sub_grp$TD2 + (sub_grp$TD2|sub_grp$NUM_PAT) , data=sub_grp)
m14 <- lmer(sub_grp$CD4 ~ sub_grp$TD + sub_grp$TD2 + (sub_grp$TD2|sub_grp$NUM_PAT) , data=sub_grp)
m15 <- lmer(sub_grp$CD4 ~ 1 + sub_grp$TD2:sub_grp$TD + (sub_grp$TD2|sub_grp$NUM_PAT) , data=sub_grp)
m16 <- lmer(sub_grp$CD4 ~ sub_grp$TD + sub_grp$TD2:sub_grp$TD + (sub_grp$TD2|sub_grp$NUM_PAT) , data=sub_grp)
# Comparing BIC and AIC
AICTable = AIC(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11,m12,m13,m14,m15,m16)
AICTable[which(AICTable$AIC == min(AICTable$AIC)),]
```

```
BICtable<-BIC(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11,m12,m13,m14,m15,m16)
BICtable[which(BICtable$BIC==min(BICtable$BIC)),]
```

We are looking for the model with the lowest BIC, and compare scores when applying either no random effects, a random effect on the intercept, or random effects on the other coefficients.

The best model is m6 which is $CD4 \sim TD + TD^2 + (1 | NUM_PAT)$ with a BIC score of 108.1738. It explains CD4 as a linear function of both TD and TD squared, with a random effect on the intercept depending on NUM_PAT.

0.4

Given that the best model out of the bunch is non reliant on dd4Tddl at all, we would argue that it does not have an effect on CD4 count.

Part 2

0.5

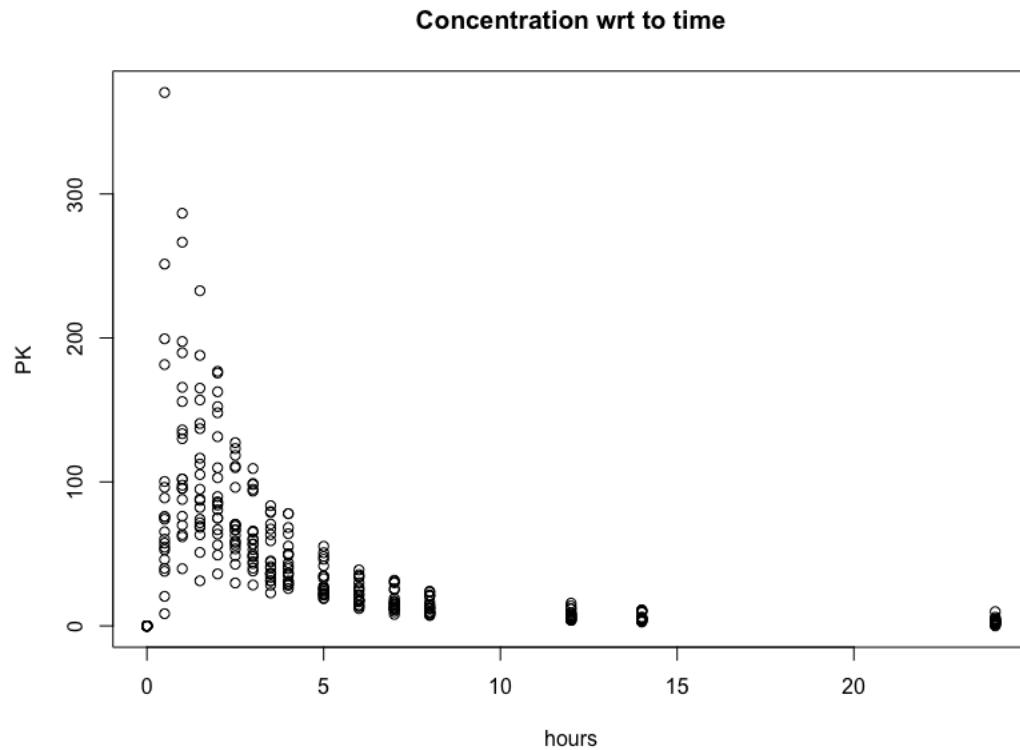


Figure 6: We plotted all datapoints of PK with respect to time in hours. It seems the PK values peak early on and exponentially decrease afterwards

0.6

We start by picking the following model from the PK library :

- Administration : oral/extravascular
- Delay : no delay
- Absorption : zero order

- Distribution : 1 compartment
- Elimination : linear

The justification for this starting model is its simplicity : it is the simplest model that could fit the data given the biological understanding of the experiment.

When running this in Monolix, we try and see whether the Observation vs Residual plot in Log Scale shows an identifiable trend. it does :

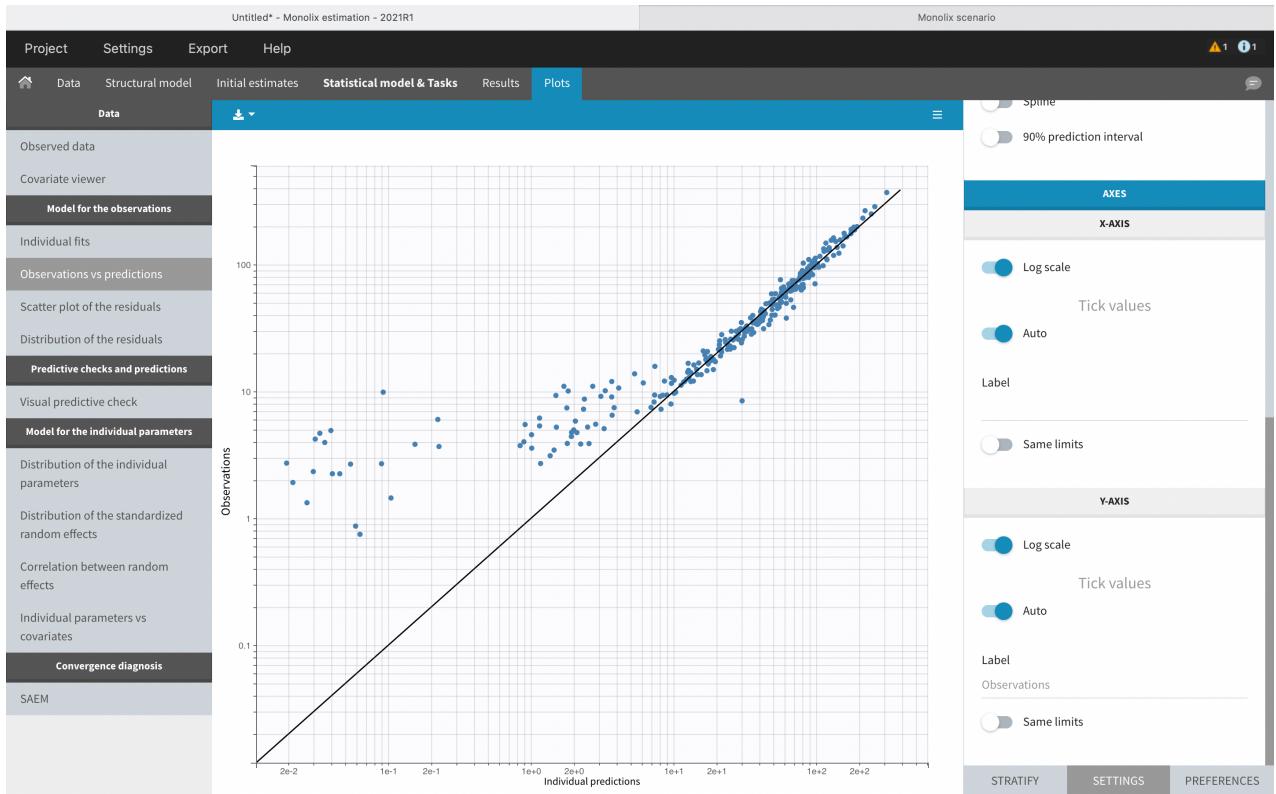


Figure 7: Clearly, the residuals have a different behavior than what is expected of them ($y=x$)

Let's look at the Individual fits, turn log scale on, and try to understand what is happening with the model :

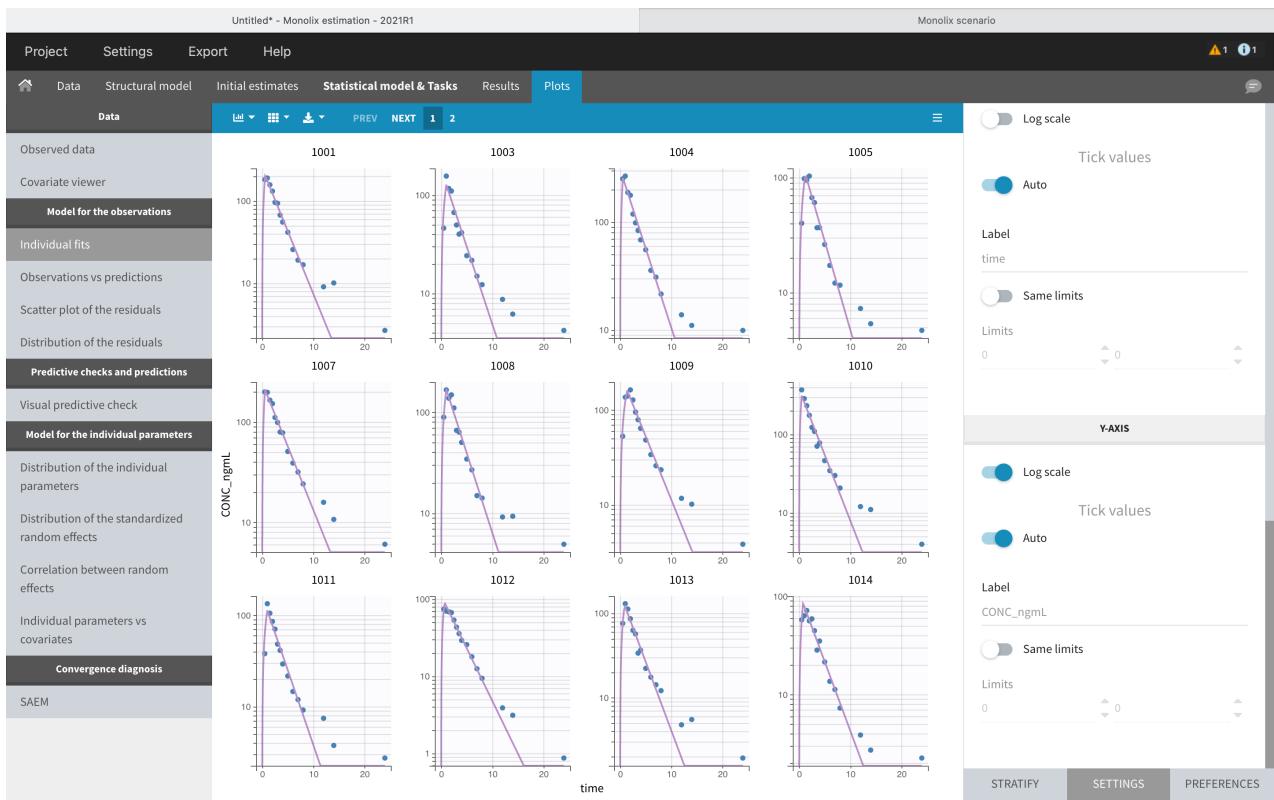


Figure 8: It seems that the elimination phase is too abrupt

To improve the elimination model, we switch to a 2 compartment model. This time the Observation vs Residual plot gives us satisfaction :

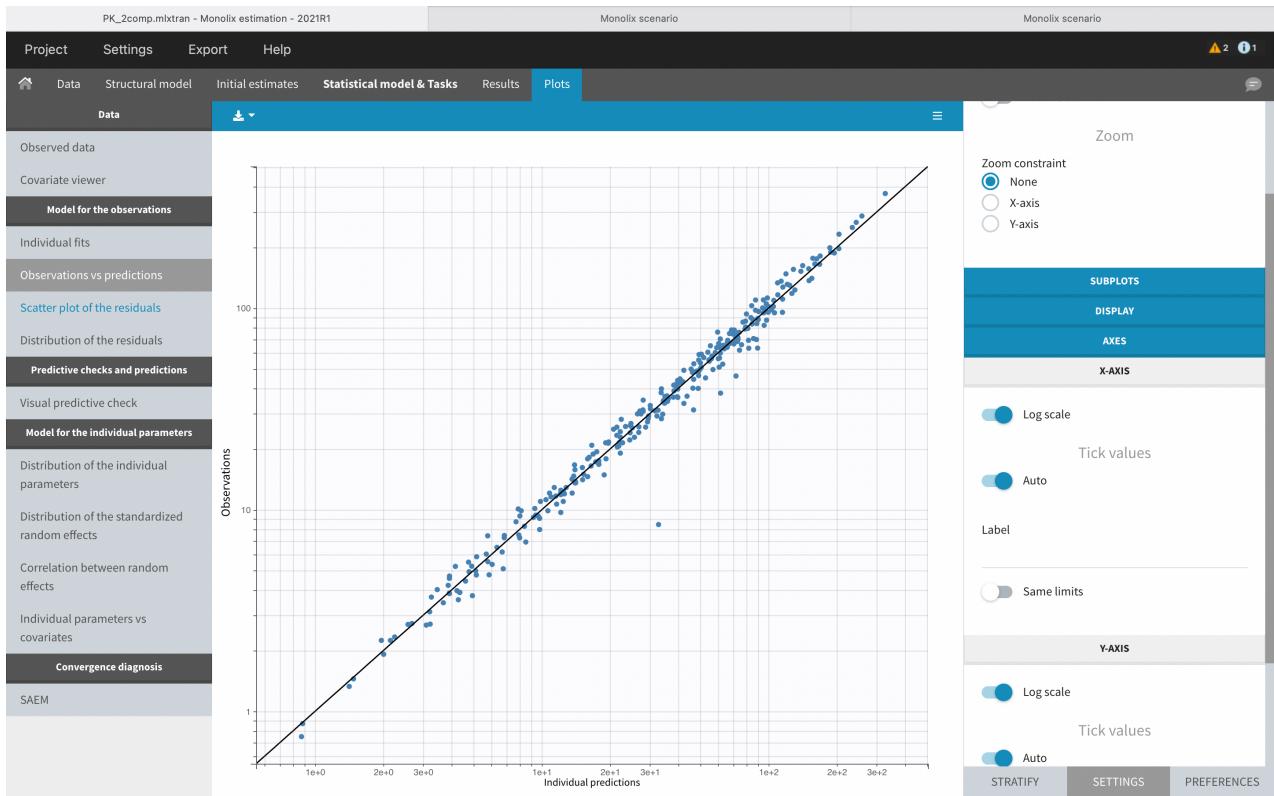


Figure 9: The Log Scale shows that there is no specific trend in residuals

The individual fits look much improved, with a more accurate elimination phase :

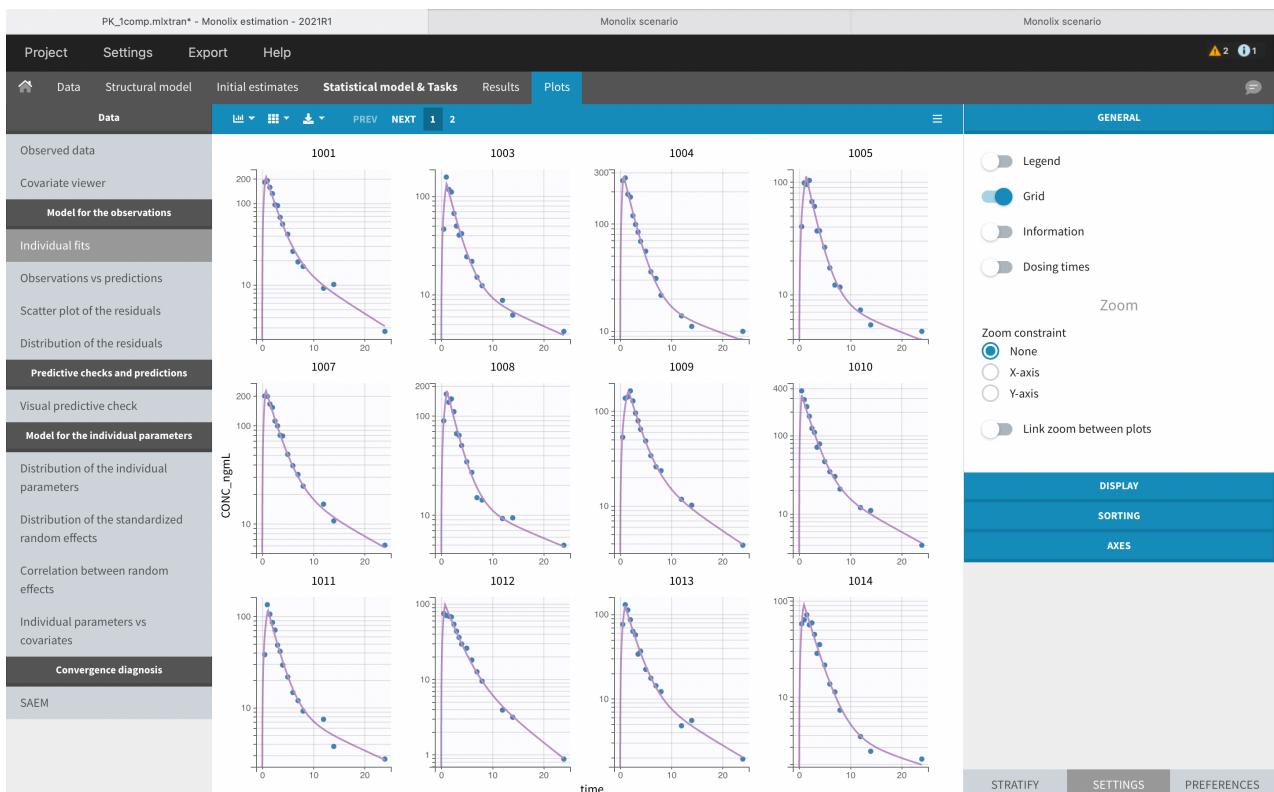


Figure 10: The curves are much more similar to the dot distribution

We will keep the 2 compartments structural model.

0.7

Let's check first the Distribution of individual parameters and the Distribution of the standardized random effects :

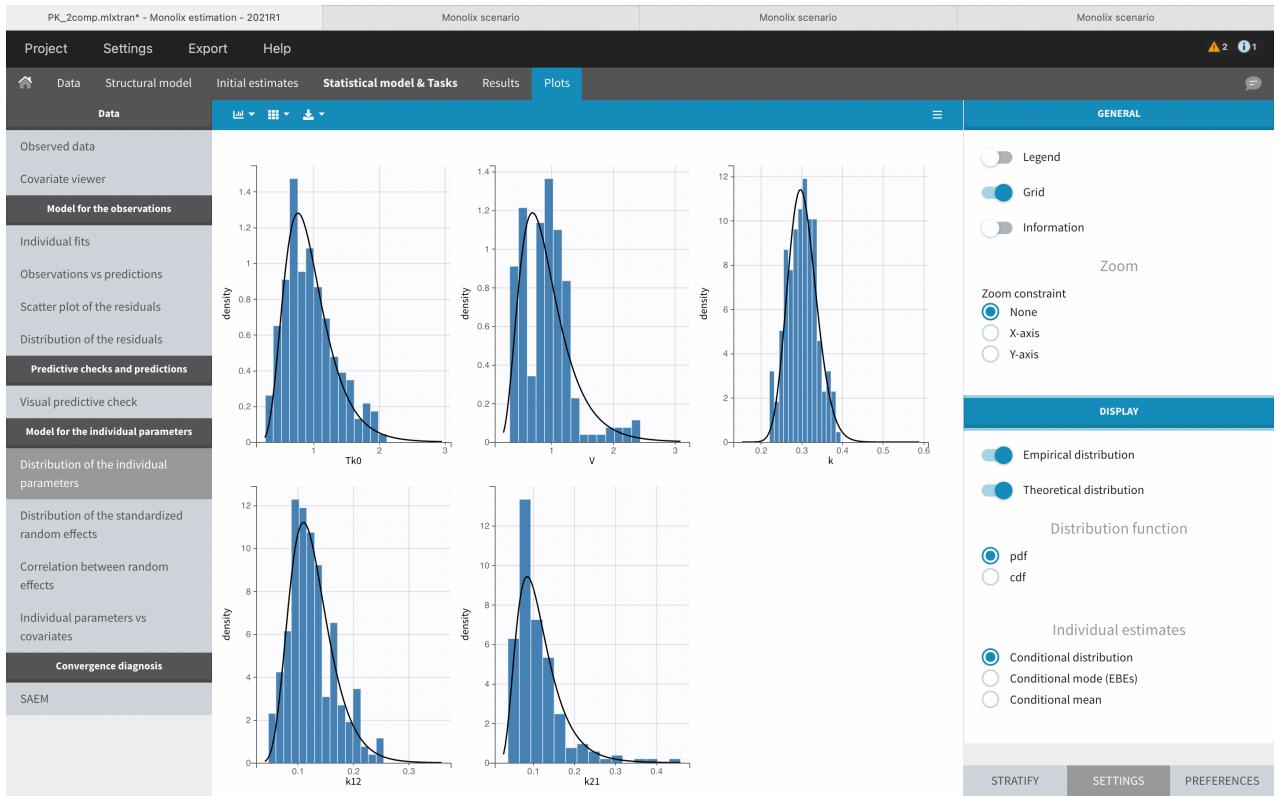


Figure 11: It looks like residuals are all normally distributed, which is reassuring

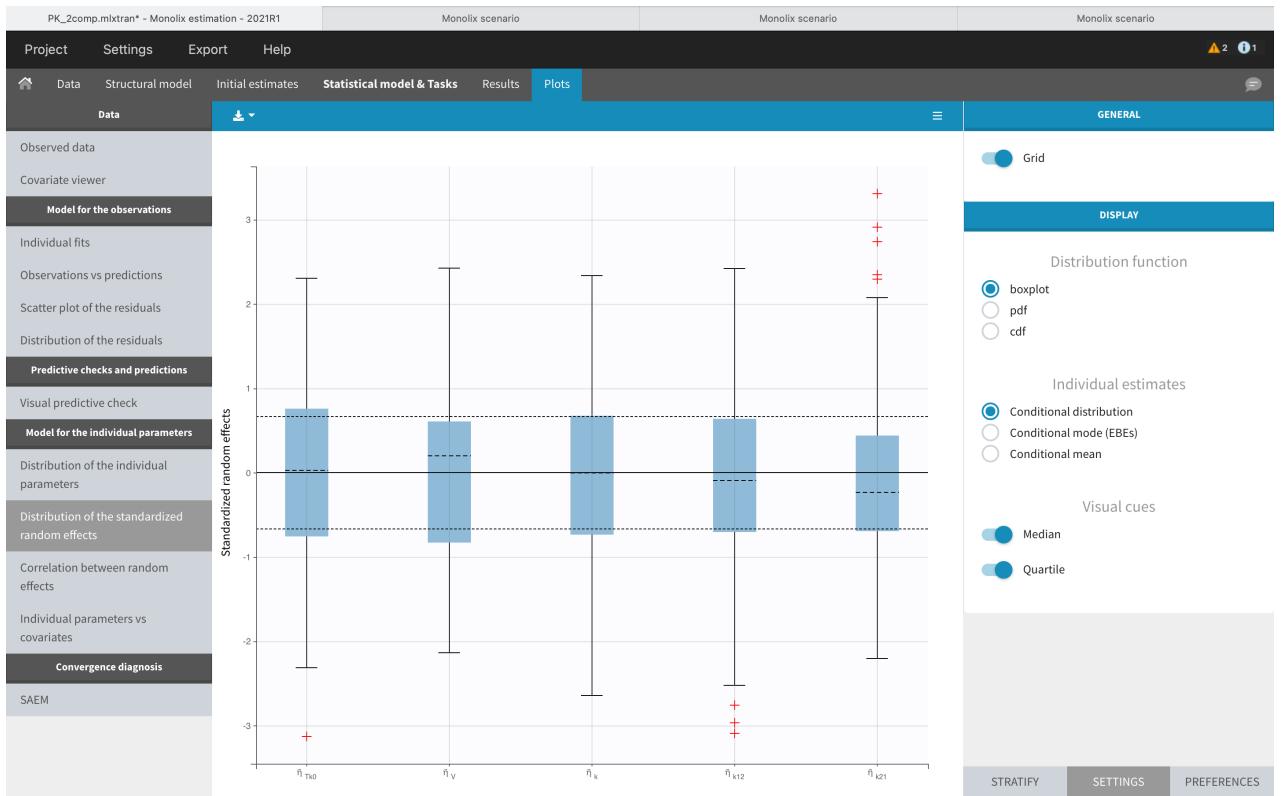


Figure 12: This graph confirms it

To find the best covariate model, we go in the results Tab, and look at low p-values. We check the covariate with the lowest one and run the model again. We also check the BIC value to make sure it decreases before going any further. Helped by the Proposal Panel, we come up with the following model :

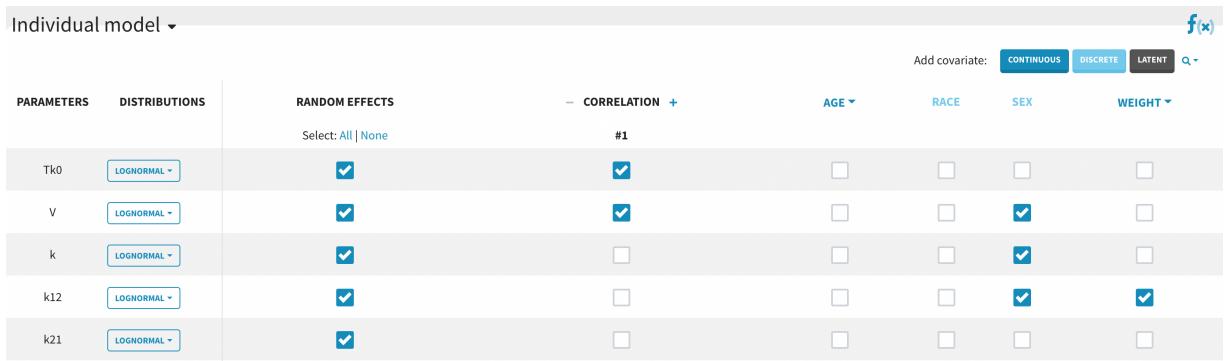


Figure 13: except the Correlation ticks, the BIC score is greatly improved with these covariates

0.8

The best model for random effects is built with the Random Effects pane in the Test tabs. We use the suggested model and obtain a BIC score of 595 :

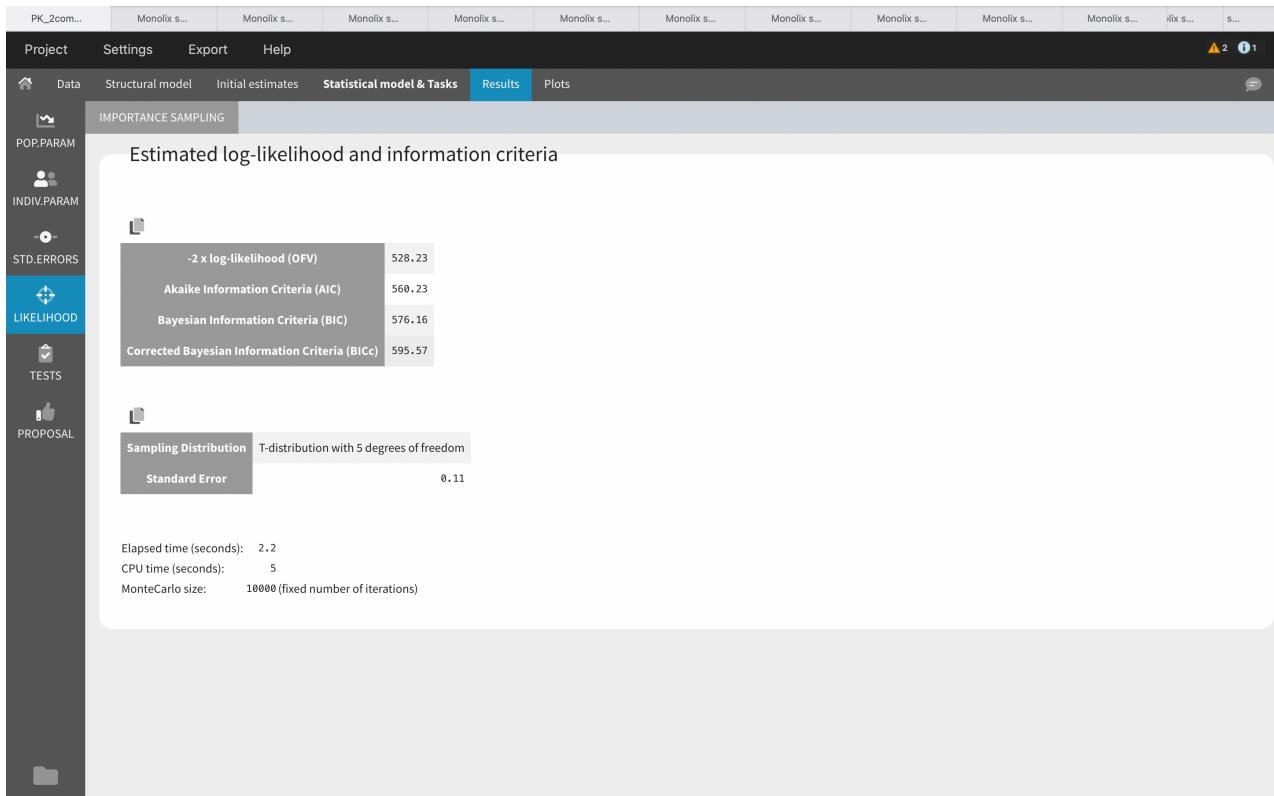


Figure 14: After having looked at Results, we removed k21 from the random effects model

We observe in the Pop Parameters pane in the Results Tab; that the value for a is 0. We change the Error Model from Combined to Proportional, and observe a drop of the BIC score, so that we validate this new model :

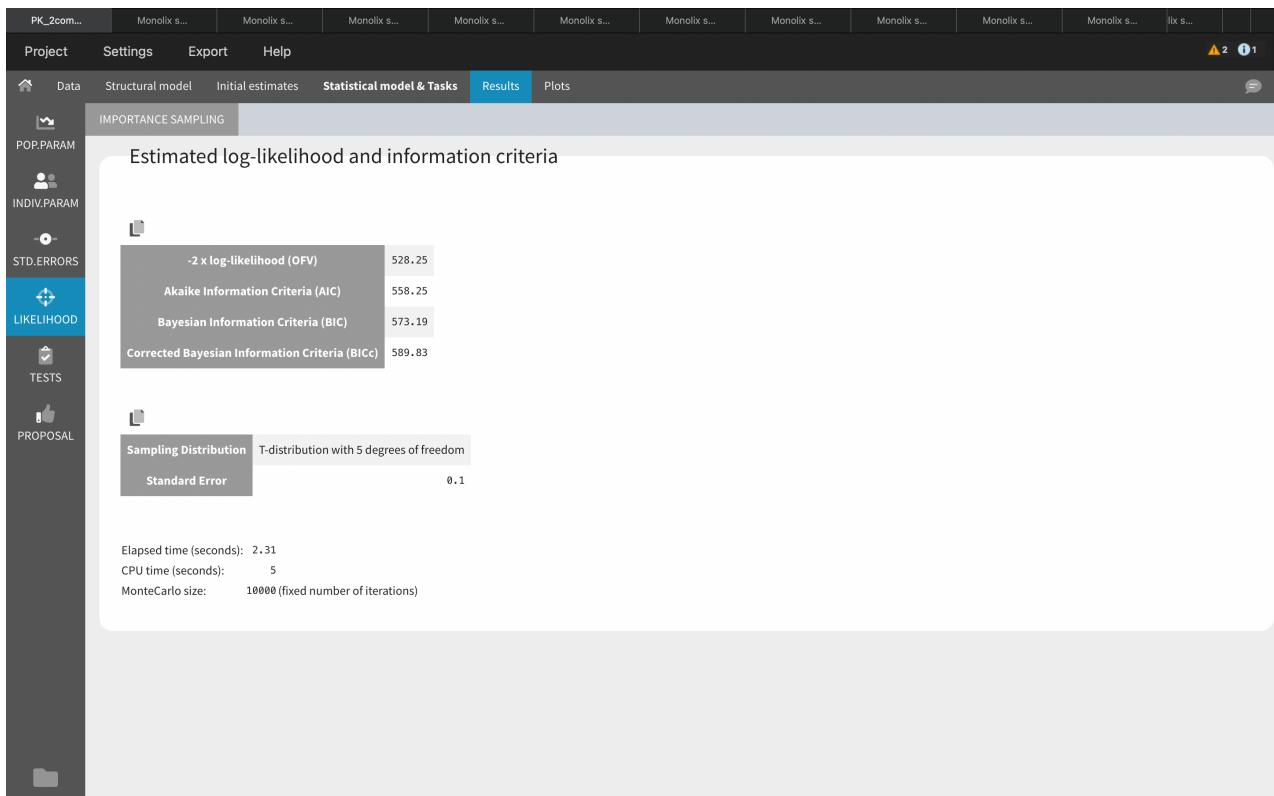


Figure 15: 589 is the best BIC score