Analysis of Anti-cancer Effect of NCTD with LMM and NLMM

Hsiao-Chieh Wu, Kasumi Tsai, and Bin Fang
Department of Mathematics &Statistics, San Diego State University, San Diego, USA

Abstract

To study the anti-cancer effect of NCTD, we fit the longitudinal data of cell proliferation with linear mixed models (LMMs) and nonlinear mixed models (NLMMs). Using top-down and step-up strategy of model selection, we find the "best" LMM and NLMM respectively, based on LRT and information criteria. The "best" LMM is considered as the final "best" model, because it has smaller AIC and BIC than the "best" NLMM and its residual pattern is better. By analysis of the estimated parameters in the final model, we confirm that NCTD inhibits the proliferation of the cancer cells.

Introduction

Cancer is one of the major health issues people have. It remains the second most common cause of death in the United States. It is estimated that 1,665,540 new cancer cases are to be diagnosed and 585,720 cancer deaths in the Unites States in 2014. NCTD (Norcantharidin) is an effective anti-cancer drug with a relatively small side effect, and it has been used in clinical cases right now.

Here, the data set we used is from a longitudinal biological experiment concerning the anti-cancer effect of a drug NCTD. Seven different concentrations (0, 5, 10, 20, 40, 80, and 120 μg/mL) of NCTD were added to the liquid media, and the concentration of cancer cells were measured at five different times (0, 12, 24, 36, 48 hours) after incubation. By comparing the concentration of cancer cells in different media and fitting them with reasonable statistical models, we may find how NCTD affect the proliferation of cancer cells.

We have 35 observations (35 measures of the concentration of cancer cells in different cases) in total. The model we use to fit the data includes linear mixed models (LMMs) and an exponential growth curve models, which is one type of nonlinear mixed models (NLMMs). By model selection, a LMM is found the "best" model, and then the parameters and diagnostics of this model are analyzed.

Methodology

1. Our data and models

The data set can be found in Appendix A and its hierarchical structure is attached in Appendix B. The response variable in this study is the concentration of cancer cells (noted as "cell" in our model). The first level covariate is the time variable (noted as "time" in our model) and the second level covariate is the concentration of NCTD in the cell media (noted as "drug" in our model). The variable cell, time and drug are all continuous variables. Additionally, we make a new variable "drug.f" just for grouping the data by treating "drug" as a dummy variable.

In Appendix C, the descriptive statistical analysis of the data is attached. It seems the pattern of the proliferation of cancer cells is not a linear relationship with the time, but a parabola-like curve. Therefore we consider including the both time and time square in the LMM. And due to the object of our study, the term "drug" is supposed to be in our model to analysis the anti-cancer effect of NCTD under different concentrations. Thus, in the LMM, there are at least three fixed effects - time, time square and drug, and two time-correlated random effect - time and time square. We assume that the random effects are random sample from normal distribution and there are no association between random effects and residuals.

In an ideal environment with unlimited nutrition and space, the number of cells would possibly increase as an exponential curve. Although this is unrealistic in the experimental environment since the nutrition and space are limited, we try to fit the data with exponential mixed models and analysis if this model is appropriate for the data. In the exponential mixed models cell~N*(a+b*drug)^time, there are three fixed effects and three random effects, of which N indicates the initial amounts of cancer cells, a indicates the natural proliferation rate of cancer cells and b indicates the effect of inhibition from NCTD. Similar to the LMM, in NLMM we also assume random effects are random sample from normal distribution and there are no association between random effects and residuals.

2. Overview of the data analysis

The procedures of model selection for LMMs and NLMMs can be found in Appendix D. Both top-down and step-up strategy are used to find the "best" model, and the results are same. In the following analysis we only show the results of model selection by top-down strategy, and the codes of the step-up strategy are attached in Appendix I.

According to the top-down strategy, we start with a loaded mean structure for the model. This step involves adding the fixed effects of as many covariates (time, drug and their square terms) as possible to the model, in order to make sure that the systematic variation of the concentrations of cells has been well explained before investigating various covariance structures to describe random variation in the data. The second step is to select a covariance structure for residuals in the model, to include the remaining variation into the residuals. We will analysis if the residual variances among groups are the same or there is a heterogeneous structure. The final step is to determine if certain fix-effect parameters can be dropped from the model.

After selecting the "best" LMM and the "best" NLMM respectively, information criteria and diagnostics of the models are used to determine which one is even better. By analyzing the estimated parameters in the "best" model, we will see how NCTD effect the proliferation of cancer cells.

The data analysis is run by R software, and its codes are attached in Appendix J.

Results

Here and in the appendix we focus on the procedures and results of top-down strategy, and the final results of step-up strategy is the same. The codes of both strategies are attached in Appendix J.

1. Model selection of LMMs

1.1 random effects:

First we try to fit the model with all four fixed effects and three random effects, but the model

cannot be created since the estimation algorithm does not converge to a solution for parameter estimates. Then we delete the random intercept, and fit a new model (model.lme.2). Deleting the random intercept is reasonable in this study: the concentration of cancer cells at 0 hours (before all the cells are allocated to each media) should have been the same since they come from the same pre-incubated media. To test if time square can be dropped from model.lme.2, model.lme.3 is fit with time square. Due to the small p-value of LRT ($p<0.0001$), AIC and BIC values, we reject the null hypothesis and keep time square in the random effects.

1.2 residual structures

Second, in order to test if there is a heterogeneous structure for the residuals, a new model-model.lme.2.hete-is created with weights assigned to the residual variances of each group. Due to the small p-value of LRT ($p<0.0001$), AIC and BIC values, we reject the null hypothesis and set a heterogeneous structure for the residuals.

1.3 fixed effects

The final step of LMM selection is to determine if all fixed effects should be kept in the model. Since our object is to study the inhibitory effect of NCTD to the cell proliferation, fixed effects drug and time should be kept in the model, so their square terms are tested if it is necessary to explain the variances of the response variable. The model with the time square is dropped and cannot be created since the estimation algorithm does not converge to a solution for parameter estimates. But from the summary of model.lme.2 (in Appendix F.1), the p-value for individual test of time square is 0.0106, so we decide to keep time square term in fixed effects. The model with drug square dropped is created as model.lme.5.hete.ml with ML estimation. Due to the small p-value of LRT ($p<0.0001$), AIC and BIC values, we reject the null hypothesis and keep time square in the random effects. Thus we get the "best" LMM – model.lme.2.hete. All the results of LRT can be found in Appendix E.1-E.3.

2. Model selection of NLMMs

2.1 random effects

Similar with LMM, we first fit a full model model.nlme.1 with all fixed and random effects. To test if N, a, b are necessary for random effects, we fit three reduced models (model.nlme.2, model.nlme.3, model.nlme.4) respectively. Due to the small p-value of LRT ($p<0.0001$), AIC and BIC values, we keep N and a in random effects. And since the p-value of LRT is large ($p=0.5147$) for the test of random effect b, we decide to drop this random effect, and our final model model.nlme.4 only include random effects N and a.

2.2 residual structures

To test if there is a heterogeneous structure for the residuals, model.nlme.4.hete is created with weights assigned to the residual variances of each group. Due to the small p-value of LRT ($p<0.0001$), AIC and BIC values, we reject the null hypothesis and set a heterogeneous structure for the residuals.

2.3 fixed effects

We keep all fixed effects in NLMM without testing them, since all of them are necessary for our study ("N" indicates the initial amounts of cancer cells, "a" indicates the natural proliferation rate of cancer cells, and "b" indicates the effect of inhibition from NCTD). Thus we get the

“best” NLMM – model.nlme.4.hete. All the results of LRT can be found in Appendix G.1-G.2.

3. Diagnosis of the “best” LMM and the “best” NLMM

Residuals vs. Fitted Value Plot, QQ Plot, Estimated Density and Predicted Cell Proliferation Trajectories can be found in Appendix F.2-F.5 for the best “LMM” and Appendix H.2-H.5 for the “best” NLMM.

The Residuals vs. Fitted Value Plots are good for both models, indicating the expectation of residuals is 0 and the variance is constant, satisfying the assumption of the model. Notice that the variances of residuals are a little greater in the “best” NLMM than in the “best” LMM, indicating more variation of the response variable is explained by the LMM.

The heavy tails in the QQ Plot and the shape of Estimated Density indicates the residuals are approximately normal around the median of the fitted values but obvious far from normal at extreme small or big fitted values. This is not consistent with our assumption of the model and is a key point to improve in the future analysis.

Predicted Cell Proliferation Trajectories seems pretty good for the “best” LMM, with only one possible outlier-observation of cancer cells in media of 5μg/ml NCTD at 12 hours. Predicted Cell Proliferation Trajectories for the “best” NLMM is not as good as the “best” LMM: the trajectories do not fit well with the observation of cancer cells in media of 0, 5 and 10μg/ml NCTD.

Based on all diagnostics plots above, we can consider the “best” LMM is even better than the “best” NLMM. And this is confirmed again by AIC and BIC value attached in Appendix I: the AIC and BIC value of the “best” LMM are smaller. Therefore we choose the “best” LMM as the final “best” model.

4. Interpretation of parameter estimates in the “best” LMM

In Appendix F.1 the summary of the “best” LMM is attached. The estimated regression parameters for time and its square term is -4.975 and 0.226 respectively. So the growth curve of cancer cell proliferation decreases slightly at the beginning and increases after the vertex. A possible explanation of this pattern is as follows. When cells are incubated in a new media (at the 0 time point in this experiment) they need some time to get used to the new environment. With this, its proliferation will delay at the beginning, and some cells may even die because of unpredictable manipulating reasons. After going through this “tough” time, the cancer cells will proliferate faster and faster, like a parabola curve.

The estimated regression parameters of drug and its square term are -0.446 and 0.00227 respectively. These values indicate the inhibitory effect of NCTD to the proliferation of cancer cells. This inhibitory effect is not a simple linear relationship with the concentration of NCTD: it is a parabola-like relationship. Note the vertex of this parabola is -0.00227/ [2*(-0.446)] = 0.0025 μg/ml, which is a very low concentration of NCTD. Thus, actually in most drug concentrations (which are significantly greater than 0.0025 μg/ml), the inhibitory effects are obvious and increase as the concentration of NCTD rises.

The estimated standard deviations for the random effect time and its square term are 5.214, 0.225 respectively. And since we fit a model with heterogeneous residual variances, the estimated residual standard deviations are different among groups: its value is 1.785 for the group with NCTD concentration of 120μg/ml (set as control group in the R function), and for the other group its value is 1.785 times the multiplier (that is 1.71，3.58，6.61，11.18，21.19，19.97 respectively).

Conclusion

After the procedures of model selection, LMM Model.2.hete is considered to be the "best" model. In this model, the fixed effect includes time, drug and their square terms, and the random effects include time and its square term, and the residuals have a heterogeneous structure. The model is as follows:

$$Cell_{ti} = 257.24 - 4.975Time_{ti} + 0.226(Time\hat{}2)_{ti} - 0.446Drug_i + 0.00227(Drug\hat{}2)_i + u_{1i}Time_{ti} + u_{2i}(Time\hat{}2)_{ti} + \varepsilon_{ti}$$

Here, t denotes the time of t hours, and i denotes the ith liquid medium (or ith concentration gradient of NCTD). $\mathbf{u} = [u_{1i}, u_{2i}]' \sim N(\mathbf{0},\mathbf{D})$, $\boldsymbol{\varepsilon} = [\varepsilon_{ti}] \sim N(\mathbf{0},\mathbf{G})$ and **u** and **ε** are mutually independent. This model can be used to analyze the effect of NCTD on the proliferation of cancer cells over time.

In the future, the data set could be approached in a different way. This approach is to nest an "S" curve model (e.g. logistic growth curve) indicating the inhibitory effect within the LMMs, and this idea originates from the growth cure in Appendix C. Notice the cell amounts among different groups at 48 hours: there seems no significant differences of cell amounts between the group of 0μg/ml and 5μg/ml NCTD, or 80μg/ml and 120μg/ml NCTD; but there seems significant differences of cell amounts between the group of 10μg/ml and 20μg/ml NCTD. In other words, if the concentration of NCTD is considered as x-axis and cell amounts at 48 hours is considered as y-axis, the graph will be a S shape and the most significant drug effect lie on the concentration region from 10μg/ml to 20μg/ml. Also, a larger data set with more time and concentration gradients may help improve this model.

Bibliography

1. West, B. T., Welch, K. B., and Galecki, A. T. (2007). Linear Mixed Models: A Practical Guide Using Statistical Software, Chapman Hall/CRC Press.
2. http://www.r-project.org/
3. http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2014/

Appendix

A. Dataset and Variables

c(NTCD) μg/mL	Time				
	0hr	12hr	24hr	36hr	48hr
0	236	216	244	463	925
5	236	258	281	489	895
10	236	233	243	395	657
20	236	248	273	301	379
40	236	237	255	283	312
80	236	236	239	267	304
120	236	238	242	255	278

Subject Variables
ID = Cell ID
drug = concentrations of NCTD in the media
Time-Varying Variables
time = Incubation time of cancer cells in hours (0, 12, 24, 36, 48)
cell = Cell Concentration after incubation in media with different concentrations of NCTD

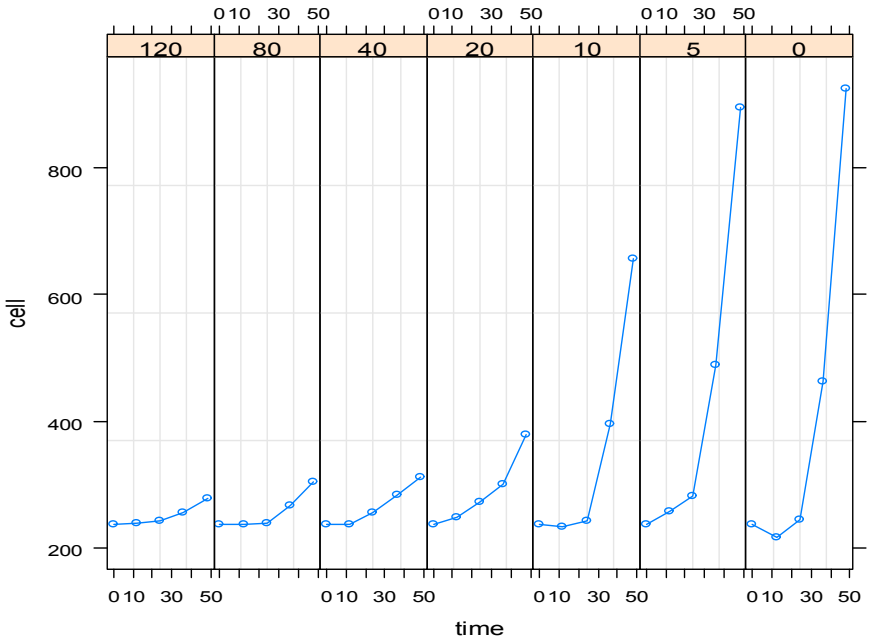
B. Hierarchical Structure of the Data

Cancer Cells (level 2)		Longitudinal Measures (level 1)	
Subject ID	Covariate	Time Variable	Dependent Variable
ID	drug	time	Cell
1	0	0	236
2	0	12	216
3	0	24	244
4	0	36	463
5	0	48	925
6	5	0	236
7	5	12	258
8	5	24	281
9	5	36	489
10	5	48	895
11	10	0	236
12	10	12	233
13	10	24	243
14	10	36	395
15	10	48	657
16	20	0	236
17	20	12	248
18	20	24	273

19	20	36	301
20	20	48	379
21	40	0	236
22	40	12	237
23	40	24	255
24	40	36	283
25	40	48	312
26	80	0	236
27	80	12	236
28	80	24	239
29	80	36	267
30	80	48	304
31	120	0	236
32	120	12	238
33	120	24	242
34	120	36	255
35	120	48	278

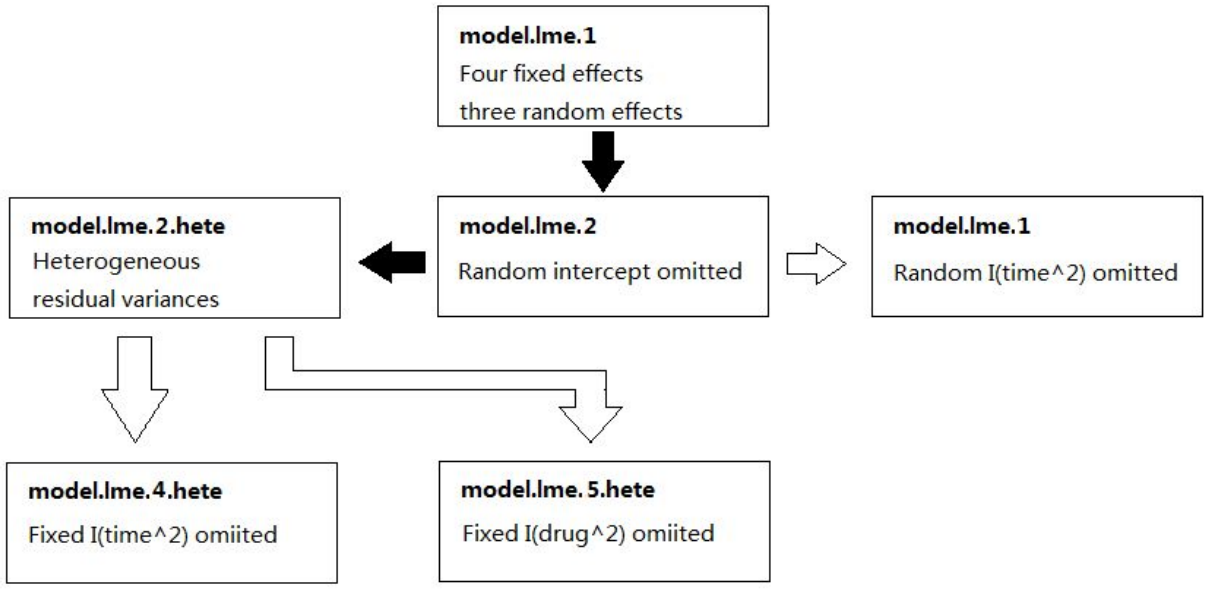
C. Summary of the data

```
> summary(NCTD)
      time          drug             cell        drug.f
 Min.   : 0   Min.   :  0.00   Min.   :216.0   120:5
 1st Qu.:12   1st Qu.:  5.00   1st Qu.:236.0   80 :5
 Median :24   Median : 20.00   Median :255.0   40 :5
 Mean   :24   Mean   : 39.29   Mean   :322.8   20 :5
 3rd Qu.:36   3rd Qu.: 80.00   3rd Qu.:302.5   10 :5
 Max.   :48   Max.   :120.00   Max.   :925.0   5  :5
                                               0  :5
```

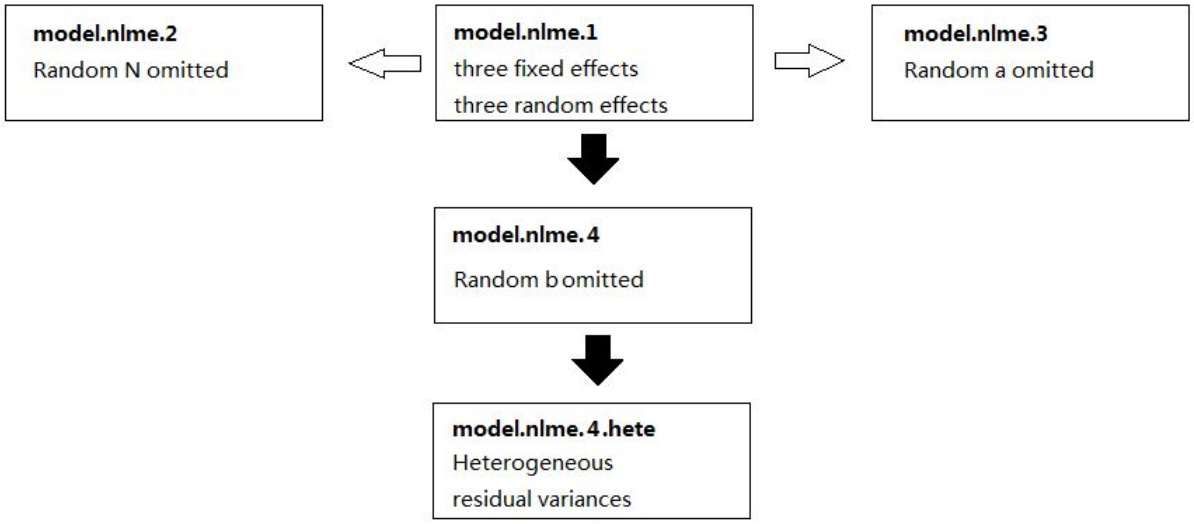


D. Overview of the model selection (Top-down Strategy)

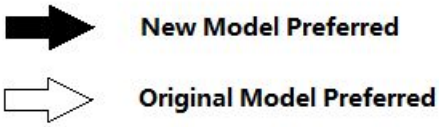
Linear Mixed Model:



Nonlinear Mixed Model:



Legend:



E. Model Selection of Linear Mixed Model (Top-down Strategy)

E.1 Step1: the structure for random effects

```
> anova(model.lme.2, model.lme.3)
             Model df      AIC      BIC    logLik   Test L.Ratio p-value
model.lme.2      1  9 377.0588 389.6696 -179.5294
model.lme.3      2  7 425.7070 435.5154 -205.8535 1 vs 2 52.6482  <.0001
```

E.1 Step2: the structure for residuals

```
> anova(model.lme.2, model.lme.2.hete)
                 Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.lme.2          1  9 377.0588 389.6696 -179.5294
model.lme.2.hete     2 15 355.8080 376.8260 -162.9040 1 vs 2 33.25079  <.0001
```

E.3 Step3: the structure for fixed effects

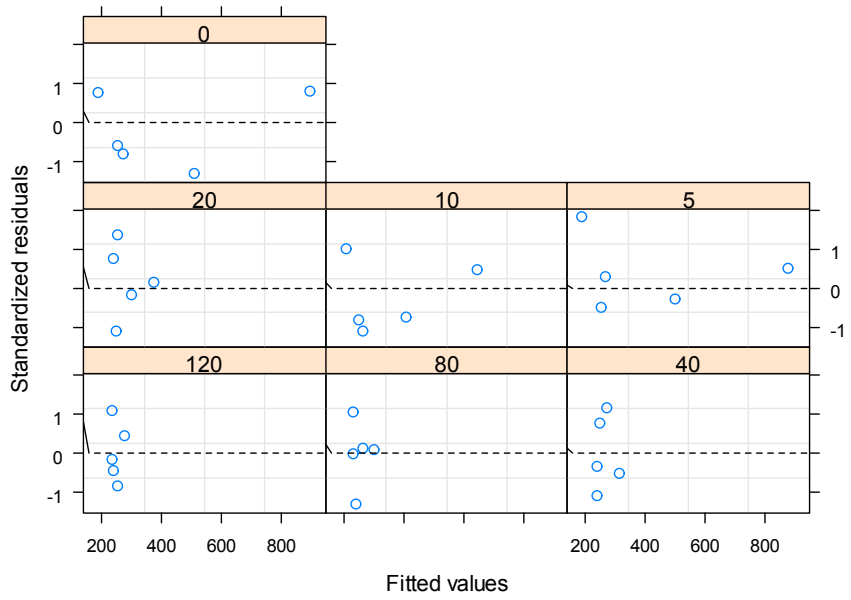
```
> anova(model.lme.2.hete.ml, model.lme.5.hete.ml)
                    Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.lme.2.hete.ml     1 15 332.7860 356.1162 -151.3930
model.lme.5.hete.ml     2 14 335.2948 357.0697 -153.6474 1 vs 2 4.508797  0.0337
```

F. Diagnostics of the “Best” Non-linear Mixed Model

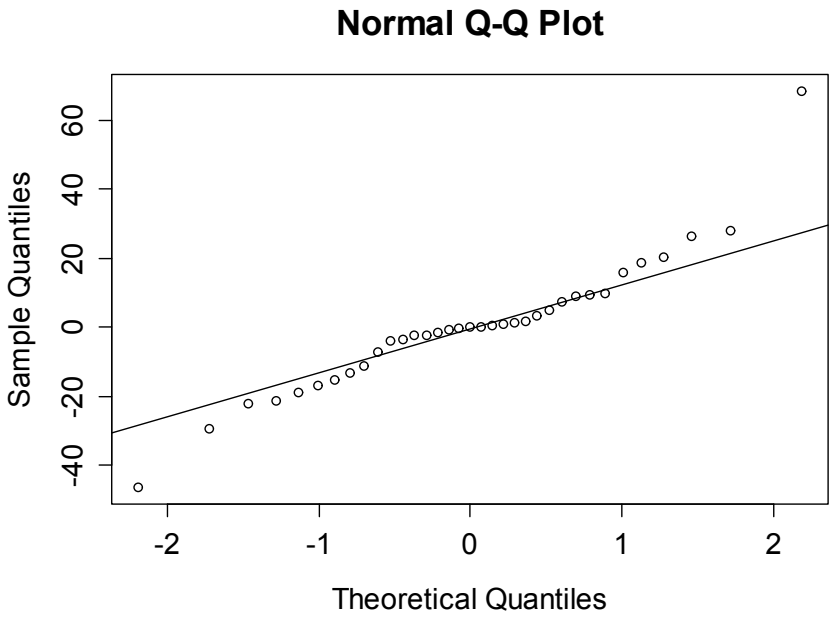
F.1 Summary of the “Best” Model

```
> summary(model.lme.2.hete)
Linear mixed-effects model fit by REML
 Data: NCTD
      AIC      BIC    logLik
  355.808 376.826 -162.904

Random effects:
 Formula: ~time + I(time^2) - 1 | drug.f
 Structure: General positive-definite, Log-Cholesky parametrization
          StdDev    Corr
time      5.2137479 time
I(time^2) 0.2248099 -1
Residual  1.7847206

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | drug.f
 Parameter estimates:
      120        80        40        20        10         5         0
 1.000000  1.712204  3.583446  6.607110 11.182108 21.188955 19.974859
Fixed effects: cell ~ time + I(time^2) + drug + I(drug^2)
               Value Std.Error DF  t-value p-value
(Intercept) 257.23810  6.285945 26 40.92274  0.0000
time         -4.97536  1.981932 26 -2.51036  0.0186
I(time^2)     0.22551  0.085289 26  2.64409  0.0137
drug         -0.44641  0.161133  4 -2.77047  0.0503
I(drug^2)     0.00227  0.000970  4  2.33835  0.0795
 Correlation:
          (Intr) time   I(t^2) drug
time       0.019
I(time^2) -0.026 -0.999
drug      -0.918 -0.028  0.028
I(drug^2)  0.841  0.025 -0.025 -0.983

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-1.3075809 -0.6812068 -0.0125343  0.7444517  1.8128981

Number of Observations: 35
Number of Groups: 7
```

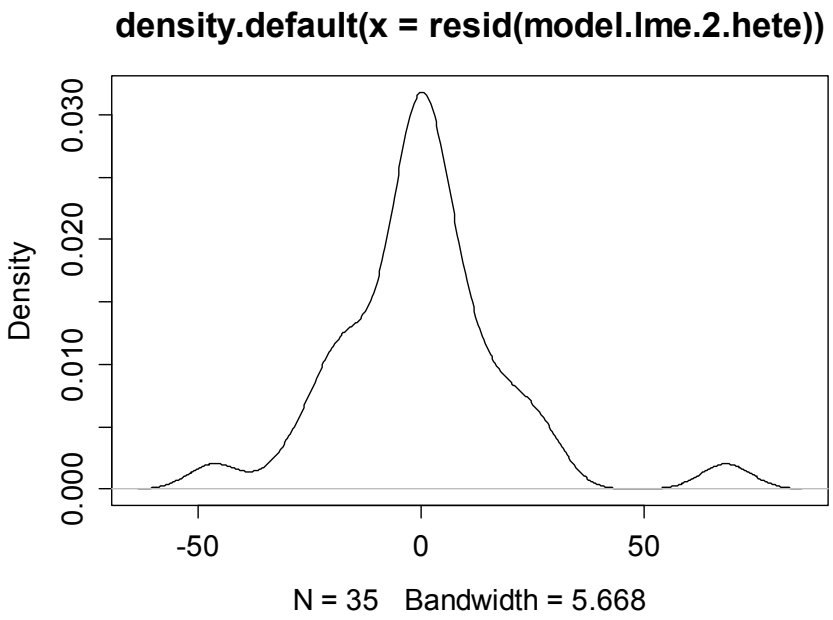
F.2 Residuals vs. Fitted Value Plot



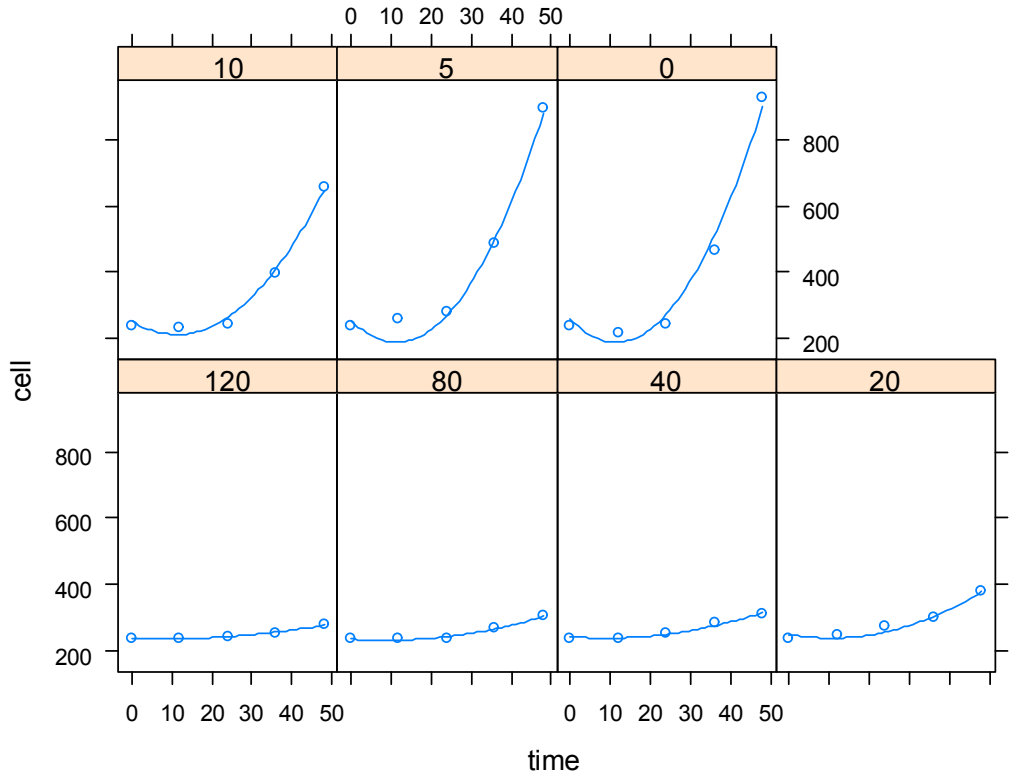
F.3 QQ Plot



F.4 Estimated Density of residuals



F.5 Predicted Cell Proliferation Trajectories for Different Concentrations of NCTD



G. Model Selection of Nonlinear Mixed Model (Top-down Strategy)

G.1 Step1: the structure for random effects

```
> anova(model.nlme.1, model.nlme.2)
             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.nlme.1     1 10 407.9084 423.4619 -193.9542
model.nlme.2     2  7 412.1363 423.0237 -199.0681 1 vs 2 10.22787  0.0167
> anova(model.nlme.1, model.nlme.3)
             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.nlme.1     1 10 407.9084 423.4619 -193.9542
model.nlme.3     2  7 412.7498 423.6372 -199.3749 1 vs 2 10.84141  0.0126
> anova(model.nlme.1, model.nlme.4)
             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.nlme.1     1 10 407.9084 423.4619 -193.9542
model.nlme.4     2  7 404.1969 415.0844 -195.0985 1 vs 2 2.288559  0.5147
```

G.2 Step2: the structure for residuals

```
> model.nlme.4.hete <- update(model.nlme.4, weights=varIdent(form=~1|drug.f) )
> anova(model.nlme.4, model.nlme.4.hete)
                 Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model.nlme.4         1  7 404.1969 415.0844 -195.0985
model.nlme.4.hete    2 13 378.2049 398.4244 -176.1024 1 vs 2 37.99207  <.0001
```

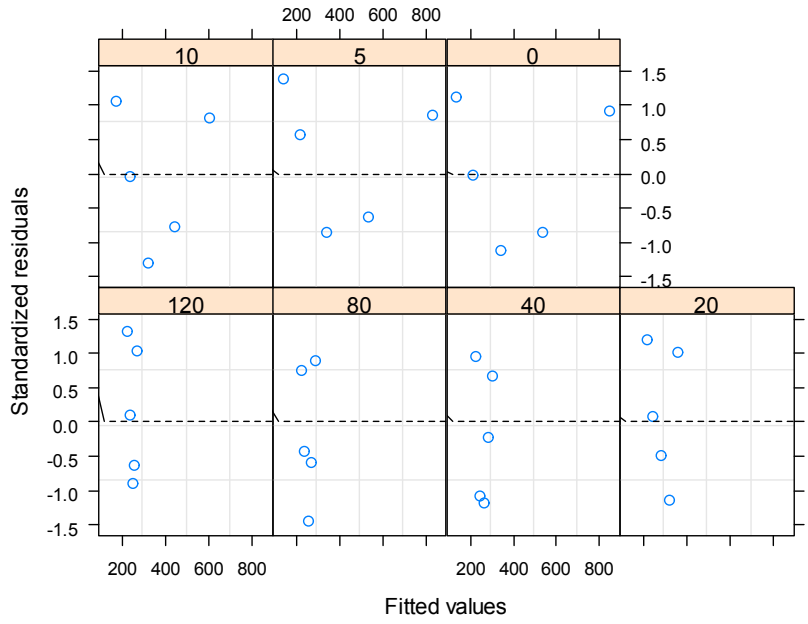
H. Diagnostics of the "Best" Non-linear Mixed Model

H.1 Summary of the "Best" Model

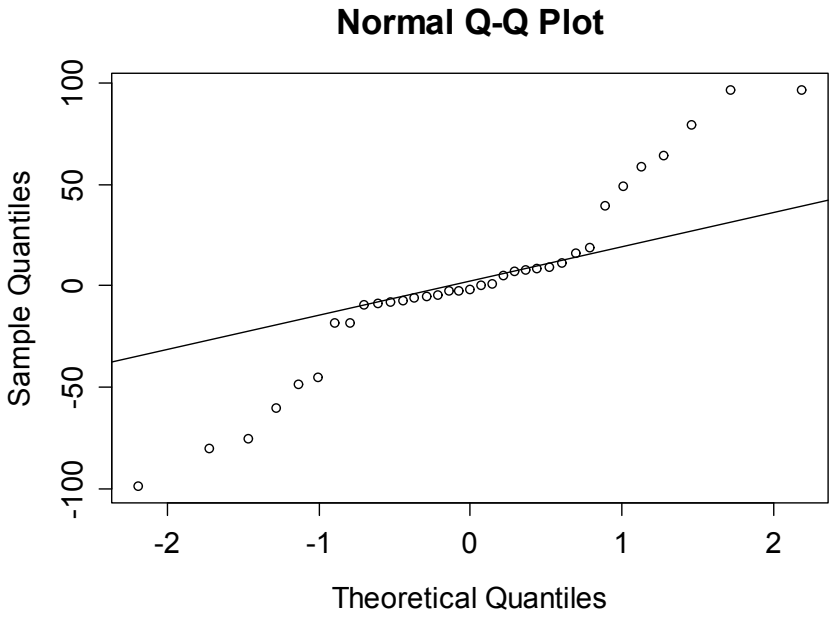
```
> summary(model.nlme.4.hete)
Nonlinear mixed-effects model fit by maximum likelihood
  Model: cell ~ N * (a + b * drug)^time
 Data: NCTD
       AIC      BIC    logLik
  378.2049 398.4244 -176.1024

Random effects:
 Formula: list(N ~ 1, a ~ 1)
 Level: drug.f
 Structure: General positive-definite, Log-Cholesky parametrization
         StdDev      Corr
N        38.96579082 N
a         0.01340639 -1
Residual  6.75859324

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | drug.f
 Parameter estimates:
      120        80        40        20        10         5         0
 1.000000  1.880310  1.195799  2.345710  9.096946 10.417921 12.946276
Fixed effects: N + a + b ~ 1
      Value Std.Error DF   t-value p-value
N 192.80870 16.412561 26  11.74763   0.000
a   1.01984  0.005399 26 188.89663   0.000
b  -0.00004  0.000019 26  -1.86863   0.073
 Correlation:
  N      a
a -0.955
b -0.209 -0.052
```

```
Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-1.45505348 -0.82240863 -0.02773408  0.89934928  1.37125995

Number of Observations: 35
Number of Groups: 7
```

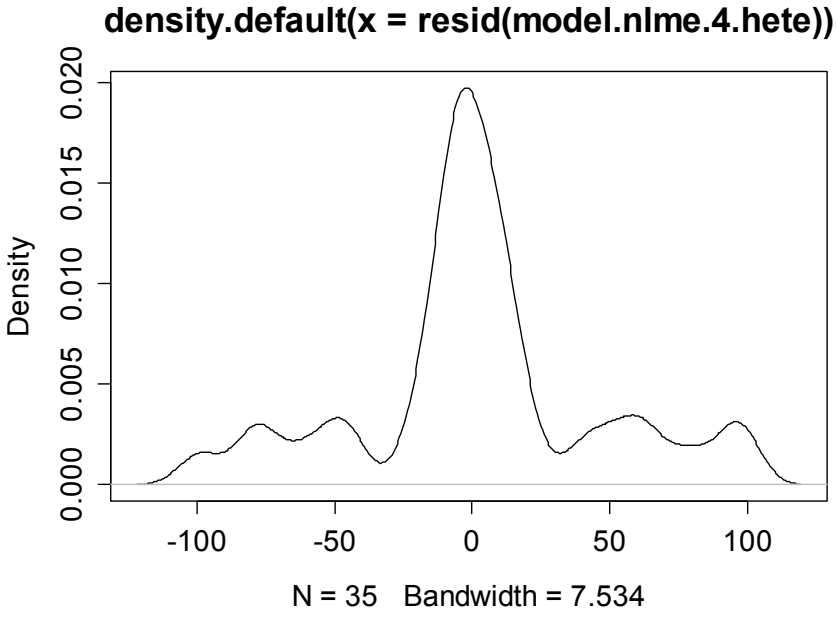
H.2 Residuals vs. Fitted Value Plot



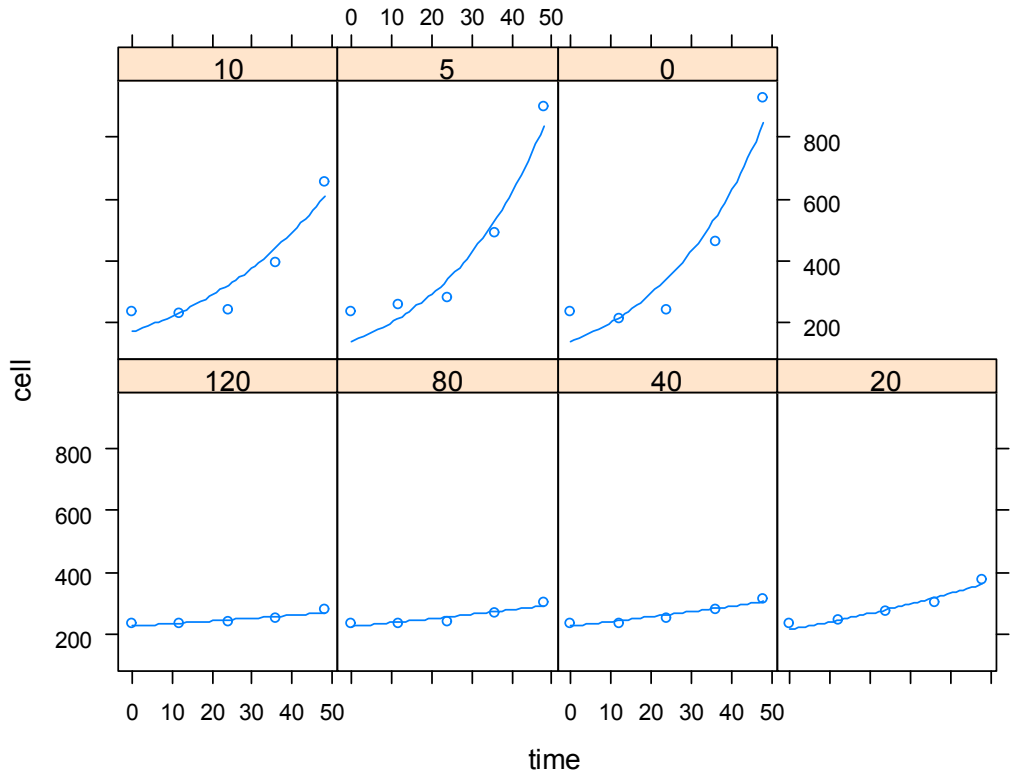
H.3 QQ Plot



H.4 Estimated Density of Residuals



H.5 Predicted Cell Proliferation Trajectories for Different Concentrations of NCTD



I. AIC and BIC value of the "best" LMM and the "best" NLMM

```
> AIC(model.lme.2.hete, model.nlme.4.hete)
                   df      AIC
model.lme.2.hete  15 355.8080
model.nlme.4.hete 13 378.2049

> BIC(model.lme.2.hete, model.nlme.4.hete)
                   df      BIC
model.lme.2.hete  15 376.8260
model.nlme.4.hete 13 398.4244
```

J. R codes

```
# Procedure 1: generate grouped data set and summary of the data
library(nlme)
raw_NCTD <- read.table("F:/Academics/Final Project/NCTD.txt", head=T)
raw_NCTD$drug.f <- factor(raw_NCTD$drug)
NCTD <- groupedData(cell ~ time | drug.f, data=raw_NCTD)
NCTD
summary(NCTD)
plot(NCTD, main='Growth curve of cancer cells')

# Procedure 2: lme model selection

# Top-down Strategy
# Step1: random effects
model.lme.1 <- lme(cell~time+I(time^2)+drug+I(drug^2), random=~time+I(time^2), data=NCTD)
# iteration limit reached without convergence

# We drop the random intercepts because of the experimental design
model.lme.2 <- lme(cell~time+I(time^2)+drug+I(drug^2), random=~time+I(time^2)-1, data=NCTD)

# H0: I(time^2) can be removed from random effects
model.lme.3 <- update(model.lme.2, random=~time-1)
anova(model.lme.2, model.lme.3)
# Conclusion: I(time^2) cannot be removed random effects
# model.lme.2 is the winner

# H0: Heterogeneous residual variances among groups
model.lme.2.hete <- update(model.lme.2, weights=varIdent(form=~1|drug.f) )
anova(model.lme.2, model.lme.2.hete)
# Conclusion: There are heterogeneous residual variances among groups
# model.lme.2.hete is the winner

# Step2: fixed effects
# H0: I(time^2) can be removed from fixed effects
model.lme.2.hete.ml <- update(model.lme.2.hete, method='ML')
model.lme.4.hete.ml <- update(model.lme.2.hete.ml, fixed=cell~time+drug+I(drug^2))
# iteration limit reached without convergence
# The growth curve seems like parabola, so we keep I(time^2) in fixed effects
```

```
# H0: I(drug^2) can be removed from fixed effects
model.lme.5.hete.ml <- update(model.lme.2.hete.ml, fixed=cell~time+I(time^2)+drug)
anova(model.lme.2.hete.ml, model.lme.5.hete.ml)
# Conclusion: I(drug^2) cannot be removed from fixed effects
# model.lme.2.hete.ml is the winner

# top-down lme winner: model.lme.2.hete

# Step-up Strategy
model.lme.6 <- lme(cell~1, random=~time+I(time^2)-1, data=NCTD)

# Step1: random effects
# H0: I(time^2) can be removed from random effects
model.lme.7 <- update(model.lme.6, random=~time-1)
anova(model.lme.6, model.lme.7)
# Conclusion: I(time^2) cannot be removed random effects
# model.lme.6 is the winner

# H0: Heterogeneous residual variances among groups
model.lme.6.hete <- update(model.lme.6, weights=varIdent(form=~1|drug.f) )
anova(model.lme.6, model.lme.6.hete)
# Conclusion: There are heterogeneous residual variances among groups
# model.lme.6.hete is the winner

# Step2: fixed effects
# H0: I(time^2) dose not need to be added to fixed effects
model.lme.6.hete.ml <- update(model.lme.6.hete, fixed=cell~time, method='ML')
model.lme.7.hete.ml <- update(model.lme.6.hete.ml, fixed=cell~time+I(time^2))
anova(model.lme.6.hete.ml, model.lme.7.hete.ml)
# Conclusion: I(time^2) should be added to fixed effects
# model.lme.7.hete.ml is the winner

# H0: I(drug^2) dose not need to be added to fixed effects
model.lme.8.hete.ml <- update(model.lme.7.hete.ml, fixed=cell~time+I(time^2)+drug)
model.lme.9.hete.ml <- update(model.lme.7.hete.ml, fixed=cell~time+I(time^2)+drug+I(drug^2))
anova(model.lme.8.hete.ml, model.lme.9.hete.ml)
# Conclusion: I(drug^2) should be added to fixed effects
# model.lme.9.hete.ml is the winner

# step-up lme winner: model.lme.9.hete
model.lme.9.hete <- update(model.lme.9.hete.ml, method='REML')

# model.lme.8.hete are the same with model.lme.2.hete
anova(model.lme.2.hete, model.lme.9.hete)

# model.lme.2.hete is the final lme winner
anova(model.lme.2.hete)
summary(model.lme.2.hete)
plot(model.lme.2.hete, resid(., type='p')~fitted(.) | drug.f, abline=0, lty=2)
```

```
qqnorm(resid(model.lme.2.hete))
qqline(resid(model.lme.2.hete))
plot(density(resid(model.lme.2.hete)))
plot(augPred(model.lme.2.hete))

# model.lme.5.hete is an alternatve model
model.lme.5.hete <- update(model.lme.5.hete.ml, method='REML')
summary(model.lme.5.hete)
plot(model.lme.5.hete, resid(., type='p')~fitted(.) | drug.f, abline=0, lty=2)
qqnorm(resid(model.lme.5.hete))
qqline(resid(model.lme.5.hete))
hist(resid(model.lme.5.hete))
plot(density(resid(model.lme.5.hete)))
plot(augPred(model.lme.5.hete))

# The diagnostics plot and the fitted value plot are similar for the two models
# we finally choose model.lme.2.hete based on the background
# that the drug concentration and its cytotoxicity is usually not a simple linear relationship

# Procedure 3: nlme model selection

# model: cell~N*(a+b*drug)^time #
model.nlme.1 <- nlme(cell~N*(a+b*drug)^time, data=NCTD,
                     fixed=N+a+b~1, random=N+a+b~1, start=c(N=200, a=1.05, b=-0.0001) )

# random effects
# H0: N can be removed from random effects
model.nlme.2 <- update(model.nlme.1, random=a+b~1)
anova(model.nlme.1, model.nlme.2)
# Conclusion: N cannot be removed from random effects
# model.nlme.1 is the winner #

# H0: a can be removed from random effects
model.nlme.3 <- update(model.nlme.1, random=N+b~1)
anova(model.nlme.1, model.nlme.3)
# Conclusion: a cannot be removed from random effects
# model.nlme.1 is the winner #

# H0: b can be removed from random effects
model.nlme.4 <- update(model.nlme.1, random=N+a~1)
anova(model.nlme.1, model.nlme.4)
# Conclusion: b should be removed from random effects
# model.nlme.4 is the winner #


# H0: Heterogeneous residual variances among groups
```

```
model.nlme.4.hete <- update(model.nlme.4, weights=varIdent(form=~1|drug.f) )
anova(model.nlme.4, model.nlme.4.hete)
# Conclusion: There are heterogeneous residual variances among groups
# model.nlme.4.hete is the final winner #

summary(model.nlme.4.hete)
plot(model.nlme.4.hete, resid(., type='p')~fitted(.) | drug.f, abline=0, lty=2)
qqnorm(resid(model.nlme.4.hete))
qqline(resid(model.nlme.4.hete))
plot(density(resid(model.nlme.4.hete)))
plot(augPred(model.nlme.4.hete))

# Procedure 4: final game between model.lme.2.hete and model.nlme.1.hete2
# compare info. criteria between LMM and NLMM #
AIC(model.lme.2.hete, model.nlme.4.hete)
BIC(model.lme.2.hete, model.nlme.4.hete)

# LMM model.lme.2.hete is the final winner
```