FINAL PROJECT REPORT

Rucha Girgaonkar

Course: ST537/437

Final Project Report

Submission Date: 1st May 2020

Table of Contents

1.Introduction

Exposure to lead can produce a variety of adverse health effects in infants and children. The majority of this lead exposure currently can be attributed to the pre-1978 houses which used lead-based paint (banned after 1978). The ingestion through chips and dust during normal teething and hand-to-mouth behavior in infants residing in proximity of such paints have been identified as the main mode of transmission. The US Centers for Disease Control and Prevention (CDC) has determined that children with blood levels above 10 micrograms/deciliter (μg/dL) of whole blood are at risk of adverse health effects.

Luckily, there are so-called chelation treatments that can help a child to excrete the lead that has been
Ingested. Particularly, the researchers were interested in evaluating the effectiveness of one such chelating treatment, succimer. They conducted the following study:

1. 120 children who had alarming lead levels in their blood were selected
2. They were randomly split into 3 groups of 40 children each
3. One group at random was assigned to receive a placebo (an inactive agent with no lead lowering properties)
4. One group was assigned to receive a low dose of succimer
5. And one group was assigned to receive a high dose of succimer
6. Blood lead levels were measured at the clinic for each child at baseline (time 0), before initiation of the assigned treatments
7. Each child was to return to the clinic at weeks 2, 4, 6, and 8 to get the blood lead level measured

The main interests of the study were

1. Whether succimer, in either low- or high-dose form, is effective over eight weeks in reducing blood lead levels in this population of children.
2. Whether blood lead levels in this population are associated with the age and/or gender of the child
3. Whether the effectiveness of succimer in reducing blood lead levels is associated with either or both of these factors

To solve these questions, I analyzed the data using exploratory data analysis methods. Based on those observations and certain assumptions, different models were studied and compared. The best model that described the given dataset was chosen. The model chosen through such comparative analysis was studied for the above-stated interests.

The high dimensionality of this model was then reduced after removing insignificant terms. Using this reduced model, the mean trajectories for different groups were studied. To evaluate the appropriateness of model assumptions, relevant model diagnostics were examined.

2.Methods

2.1 Exploratory Data Analysis

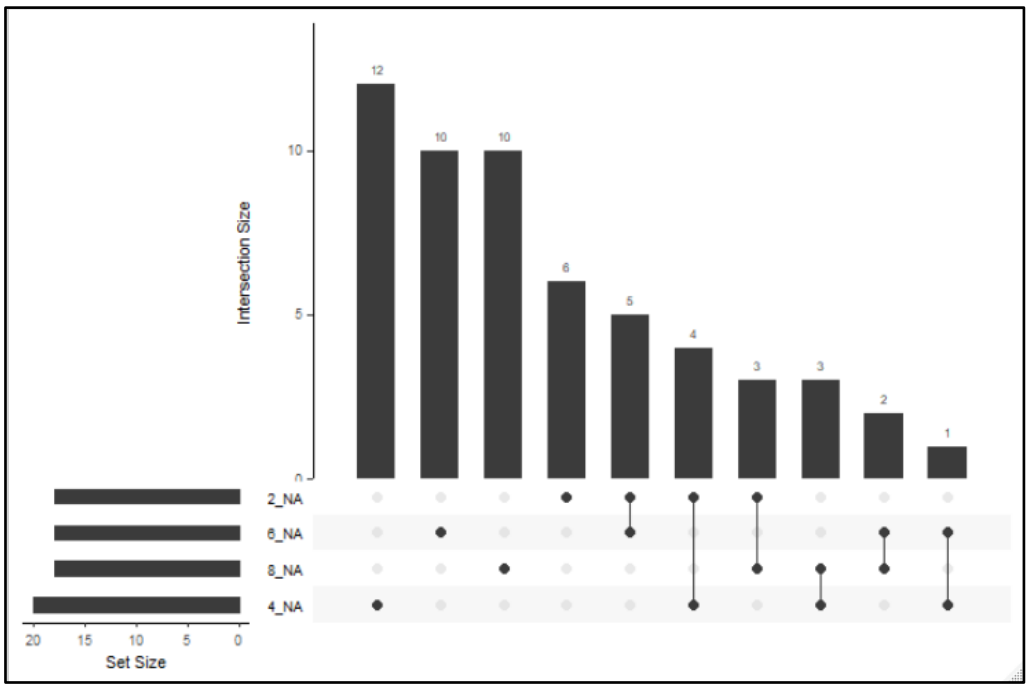
Exploratory data analysis is important to understand the variables in the given data, to clean the data and to analyze relationships between variables. Exploratory Data Analysis is a crucial step before starting the modeling of given data. It provides the context needed to develop an appropriate model – and interpret the results correctly.

The lead dataset has unbalanced data. There are missing entries in measurements of blood lead level for some weeks for some children. We should check if there is any pattern for missing values in the data. If there is some pattern of missing values, we must address/consider it in our analysis because it can have a significant effect on the conclusions that can be drawn from the data.

The details of missing values in the lead dataset are as follows:

The number of missing entries in the given data = 74.

The following graph shows the distribution of missing values across different weeks.



The summation of columns corresponding to dots in front of each week will give a total number of missing values for that week.

Weeks	No of missing entries
0	0
2	18
4	18
6	18
8	20
Total	**74**

The lines joining dots across columns are an indicator of missing values across all the weeks having dots.

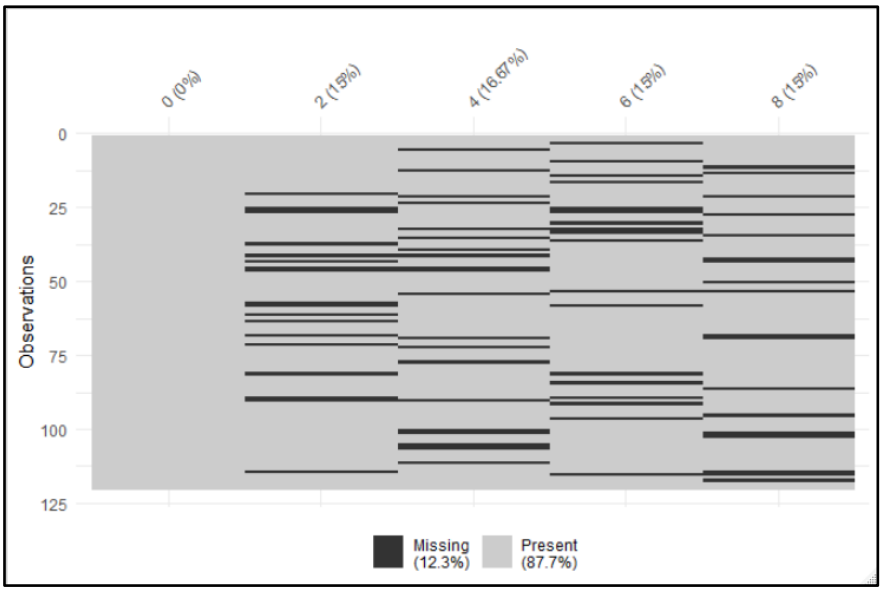
The following table shows the number of children having a particular number of missing values.

No of missing values	No of Entries
0	64
1	38
2	18
3	0
4	0
5	0

There are 64 children having measurements for all the 5 weeks. There are 38 children with the missing value of measurement for 1 week. And there are 18 children with missing values of measurements for 2 weeks.

From the following plot, we can check if there is any pattern of missing values across the dataset.

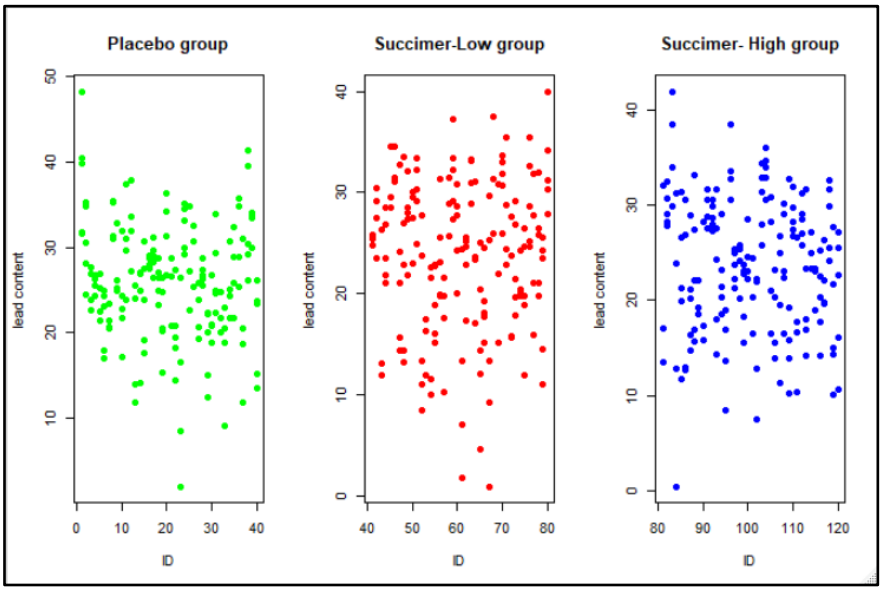
vis_miss provides an at-a-glance ggplot of the missingness inside a dataframe, coloring cells according to missingness, where black indicates a missing cell and grey indicates a present cell.



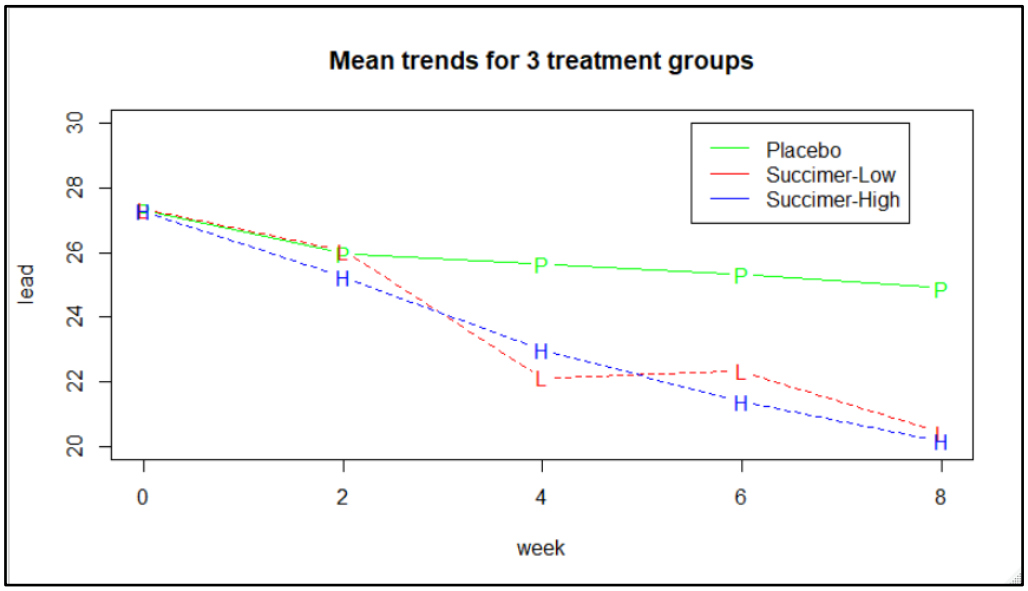
From all of the above plots and summaries, we can say that there is no particular pattern of missing values. They are randomly distributed.

Some more plots for data analysis are presented below.

Responses across three treatments.

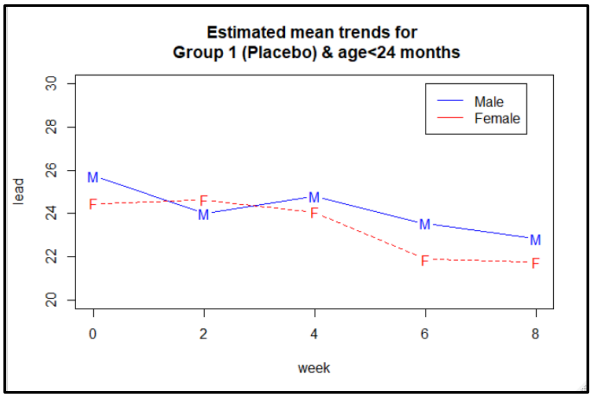


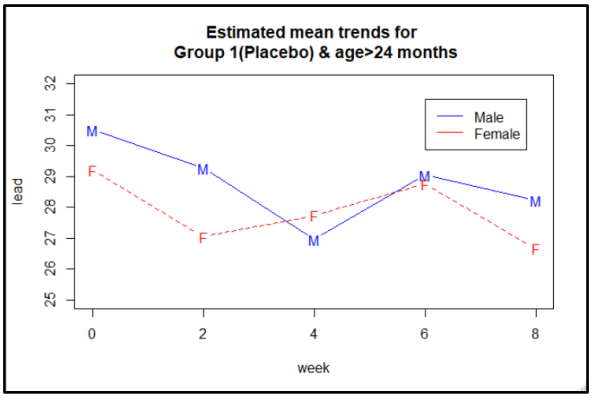
Mean trends for 3 treatment groups

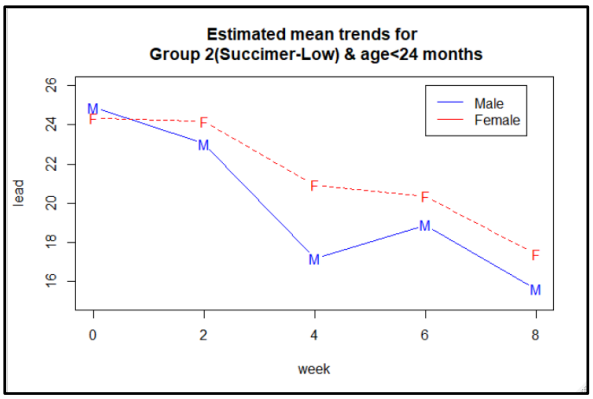


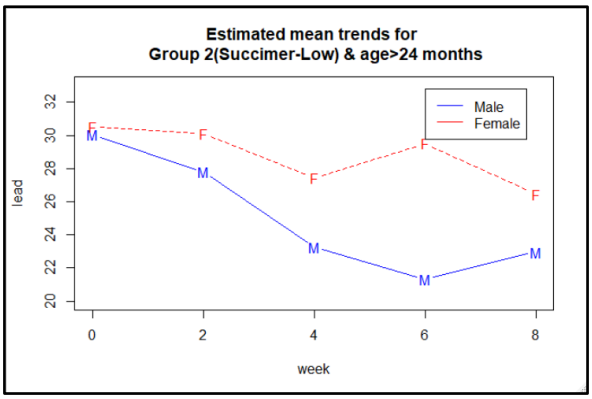
From the above plot, we can say that the mean trajectory for placebo is very much different than that of the other two treatment groups. The succimer- low and succimer- high treatment groups have a somewhat similar mean trend.

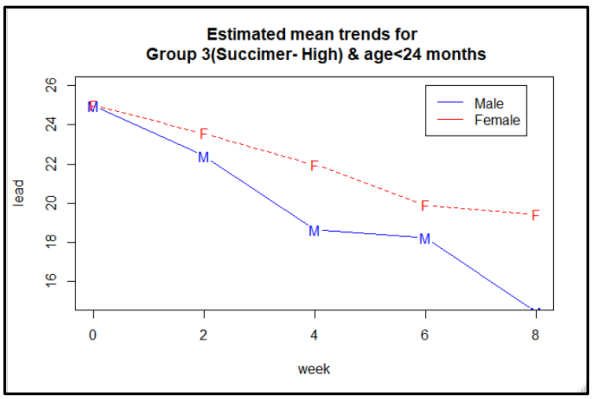
Below graphs, shows the mean trend across different treatment groups, different age groups and different genders.

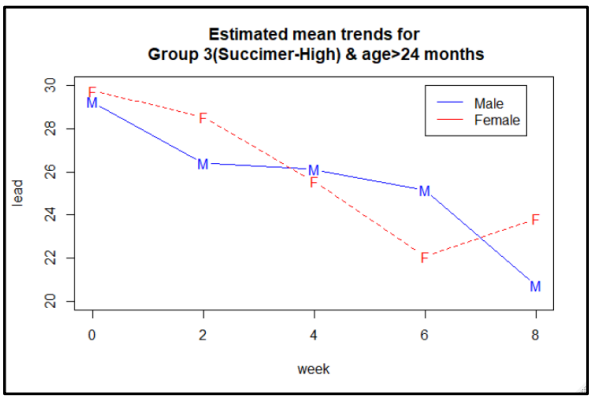












From the patterns in the above plots, we can make the following conclusions:

- For the placebo group, the blood lead level is not drastically decreasing over the weeks.
- For succimer- low and succimer- high groups, blood lead level is linearly decreasing over the weeks.
- Male and females follow an almost similar pattern.

2.2 Statistical Models

As the lead dataset is an unbalanced data, we will use a linear mixed-effects modeling approach for our analysis. This approach explicitly acknowledges two sources of variation i.e., within unit and between unit variation. Also, it allows for more general covariance structures and can accommodate additional covariate information easily.

The combined model for three treatment groups using formulation 1 can be written as follows:

$$
\begin{aligned}
Y_{ij} &= C_1\beta_{0P} + C_1\beta_{1P}week + C_1\beta_{2P}sex + C_1\beta_{3P}ind.age + C_1\beta_{4P}week*ind.age \\
&+C_1\beta_{5P}*sex*week + C_1\beta_{6P}*sex*ind.age + C_1\beta_{7P}*week*sex*ind.age \\
&+C_2\beta_{0L} + C_2\beta_{1L}week + C_2\beta_{2L}sex + C_2\beta_{3L}ind.age + C_2\beta_{4L}week*ind.age \\
&+C_2\beta_{4L}*sex*week + C_2\beta_{6L}*sex*ind.age + C_2\beta_{7L}*week*sex*ind.age \\
&+C_3\beta_{0H} + C_3\beta_{1H}week + C_3\beta_{2H}sex + C_3\beta_{3H}ind.age + C_3\beta_{4H}week*ind.age \\
&+C_3\beta_{4H}*sex*week + C_3\beta_{6H}*sex*ind.age + C_3\beta_{7H}*week*sex*ind.age + b_{0i} + b_{1i} + e_{ij}
\end{aligned}
$$

where β corresponds to fixed effects and b_{0i} and b_{1i} correspond to random effects of intercept and slope of the week respectively.

There are three treatment groups in the data. The model for an individual group can be written as follows:

For Placebo treatment group (C_1=1, C_2=0, C_3=0):

$$
\begin{aligned}
Y_{ij} &= \beta_{0P} + \beta_{1P}week + \beta_{2P}sex + \beta_{3P}ind.age + \beta_{4P}week*ind.age \\
&+\beta_{5P}*sex*week + \beta_{6P}*sex*ind.age + \beta_{7P}*week*sex*ind.age + b_{0i} + b_{1i} + e_{ij}
\end{aligned}
$$

For Succimer -Low treatment group (C_1=0, C_2=1, C_3=0):

$$
\begin{aligned}
Y_{ij} &= \beta_{0L} + \beta_{1L}week + \beta_{2L}sex + \beta_{3L}ind.age + \beta_{4L}week*ind.age \\
&+\beta_{4L}*sex*week + \beta_{6L}*sex*ind.age + \beta_{7L}*week*sex*ind.age + b_{0i} + b_{1i} + e_{ij}
\end{aligned}
$$

For Succimer -High treatment group (C_1=0, C_2=0, C_3=1):

$$
\begin{aligned}
Y_{ij} &= \beta_{0H} + \beta_{1H}week + \beta_{2H}sex + \beta_{3H}ind.age + \beta_{4H}week*ind.age \\
&+\beta_{4H}*sex*week + \beta_{6H}*sex*ind.age + \beta_{7H}*week*sex*ind.age + b_{0i} + b_{1i} + e_{ij}
\end{aligned}
$$

For the given data, it is assumed that –

The error variance-covariance structure is the same for all three treatment groups.

Random effects are present for intercept and slope of the week with the general covariance structure across three treatment groups.

The linear model for the i^{th} subject can be written as-

$Y_i = X_i\beta + e_i$, where $e_i =$ random errors

and $\mathrm{cov}(\mathrm{e}_i) = \sum_i$

'i' denotes repeated measurement on each subject.

Various models are built according to different error covariance structures. 6 of those models are as follows:

1) Independent, where error variance does not change over weeks,
2) Independent, where error variance changes over weeks,
3) AR(1) correlation structure, where error variance does not change over weeks
4) AR(1) correlation structure, where error variance changes over weeks
5) Unstructured, where error variance does not change over weeks
6) Unstructured, where error variance changes over weeks

The R-code and output of each model are given in the appendix.

The relevant random effects structure, correlation structure and weights for error variance are defined for each model.

In the unstructured model, one particular issue about using the corSymm() structure is that the covariate "week" for this correlation structure must be consecutive integers. But in given data, weeks are not consecutive and also for some individuals, some values are missing. Thus, we first re-label the time points as 1, 2,..., 7 in "timefact" variable.

1. The mean formula ("meanform") does not change – we are still using "week" as a continuous covariate
2. The newly created variable "timefact" (re-labeled time points) is only used to specify the correlation structure and weights.

2.2.1 Selection of best model using information criteria

In order to choose the best model among the above six, we will use information criteria.

Information criteria provide an informal approach to comparing competing models. They are constructed based on the idea of balancing between the goodness of fit and the number of parameters involved in the model. The larger model will give a better fit, that is, a larger likelihood. Thus, to compare the two models, we need to look at both the likelihood value and the number of parameters in the model. Thus, the information criteria are essentially penalized version of the maximized (log-)likelihood.

The two information criteria chosen to compare the above models are - Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC).

The following table will give AIC, BIC values for all models.

Sr. no	Model Description	Model name	Degrees of freedom	AIC	BIC
1	Independent (error variance does not change over weeks)	fit.in.ev	28	3079.090	3198.518
2	Independent (error variance changes over weeks)	fit.in.uv	32	3082.369	3218.859
3	AR(1) correlation structure (error variance does not change over weeks)	fit.ar1.ev	29	3081.090	3204.784
4	AR(1) correlation structure (error variance changes over weeks)	fit.ar1.uv	33	3084.369	3225.124
5	Unstructured (error variance does not change over weeks)	fit.un.ev	38	3088.428	3250.509
6	Unstructured (where error variance changes over weeks)	fit.un.uv	42	3092.238	3271.381

The log-likelihood values for the above models are as follows:

Sr. no	Model Description	Model name	log-likelihood
1	Independent (error variance does not change over weeks)	fit.in.ev	-1511.545
2	Independent (error variance changes over weeks)	fit.in.uv	-1509.185
3	AR(1) correlation structure (error variance does not change over weeks)	fit.ar1.ev	-1511.545
4	AR(1) correlation structure (error variance changes over weeks)	fit.ar1.uv	-1509.185
5	Unstructured (error variance does not change over weeks)	fit.un.ev	-1506.214
6	Unstructured (where error variance changes over weeks)	fit.un.uv	-1504.119

We prefer the model having the lowest value of AIC/BIC. From the above AIC/BIC table, we can say that fit.in.ev (Independent (error variance does not change over weeks)) is the best choice here.

2.3 Analysis of selected model and Results

The model with independent covariance structure i.e. no correlation within-subjects and errors e_i have the same variance over each week and each treatment group is selected as best fit according to lowest AIC/BIC values among all models.

Now, we will analyze this model to a greater extent.

Lets, first analyze the effect of various covariates like Gender and age on response i.e., blood lead level.

2.3.1 Association of Gender and Age with the blood lead level

We can use hypothesis testing to check the association of Gender and Age with the blood lead level. For our dataset, we will use a t-test and Wald test to examine the effects of Gender and Age.

As we are interested in certain linear combinations of components of β, we will use the L matrix specific to conditions we wanted to check.

To test the association of Gender with blood level, the L matrix can be constructed as follows:

$$\begin{bmatrix}
0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1
\end{bmatrix}$$

The β matrix will be

$$[C_1\beta_{0P}, C_2\beta_{0L}, C_3\beta_{0H}, C_1\beta_{1P}week, C_1\beta_{1P}sex, C_1\beta_{3P}ind.age, C_2\beta_{1L}week, C_2\beta_{2L}sex,$$
$$C_2\beta_{3L}ind.age, C_3\beta_{1H}week, C_3\beta_{2H}sex,\ C_3\beta_{3H}ind.age,\ C_1\beta_{4P}week*ind.age,$$
$$C_1\beta_{5P}*sex*week, C_1\beta_{6P}*sex*ind.age, C_2\beta_{4L}week*ind.age, C_2\beta_{4L}*sex*week$$
$$C_2\beta_{6L}*sex*ind.age, C_3\beta_{4H}week*ind.age, C_3\beta_{4H}*sex*week, C_3\beta_{6H}*sex*ind.age,$$
$$C_1\beta_{7P}*week*sex*ind.age, C_2\beta_{7L}*week*sex*ind.age, C_3\beta_{7H}*week*sex*ind.age]^T$$

Each row in L*β matrix will check the significance of individual elements in the β matrix.

For Gender, we have a total of 12 terms including main effects and all of the interaction effects. Hence, the L matrix has 12 rows. If we write all 1's in a single row of L, we will only get to know whether the sum of all elements is significant or not but not the significance of

the individual term. Hence, we will be checking the significance of each term through the individual row of the L matrix.

The Wald statistics and corresponding p-value are as follows:

Wald.g	p.value.g
8.23386	0.7665984

As the p-value is greater than 0.05, we will accept the null hypothesis which says $H_0 = L\beta = 0$ i.e., the terms are insignificant. Therefore, we conclude that the terms constituting the main effects and interaction effects of covariate Gender are not significant and they are not associated with blood lead level.

The results from the t-test also confirm the above analysis. All terms having gender (Main effects and interaction effect) have a p-value greater than 0.05. This indicates that these terms are not significant for the blood lead level.

Similarly, we can check the association of age covariate with the blood lead level. The L matrix to check the significance of age terms is as follows:

$$\begin{bmatrix}
0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1
\end{bmatrix}$$

The β matrix will be

$$[C_1\beta_{0P}, C_2\beta_{0L}, C_3\beta_{0H}, C_1\beta_{1P}week, C_1\beta_{1P}sex, C_1\beta_{3P}ind.age, C_2\beta_{1L}week, C_2\beta_{2L}sex,$$
$$C_2\beta_{3L}ind.age, C_3\beta_{1H}week, C_3\beta_{2H}sex,\ C_3\beta_{3H}ind.age,\ C_1\beta_{4P}week*ind.age,$$
$$C_1\beta_{5P}*sex*week, C_1\beta_{6P}*sex*ind.age, C_2\beta_{4L}week*ind.age, C_2\beta_{4L}*sex*week$$
$$C_2\beta_{6L}*sex*ind.age, C_3\beta_{4H}week*ind.age, C_3\beta_{4H}*sex*week, C_3\beta_{6H}*sex*ind.age,$$
$$C_1\beta_{7P}*week*sex*ind.age, C_2\beta_{7L}*week*sex*ind.age, C_3\beta_{7H}*week*sex*ind.age]^T$$

The Wald statistics and corresponding p-value are as follows:

Wald.a	p.value.a
98.079	1.322e-15

As the p-value is less than 0.05, we fail to accept the null hypothesis which says $H_0 = L\beta = 0$ i.e., the terms are insignificant. Therefore, we conclude that the terms constituting the main effects and interaction effects of covariate age are significant and they are associated with blood lead level.

The results from the t-test also confirm the above analysis. The main effects of covariate age have a p-value less than 0.05. This indicates that these terms are significant for the blood lead level.

2.3.2 Smaller/Reduced Model

In the previous section, we checked the association of covariates gender and age with the response blood lead level. From the analysis and results of that section, we can conclude that Gender does not have any association with blood lead level but Age (indicator age) does have an association with the response blood lead level.

We also observed that the interaction terms for week and indicator age are not significant.

After removing insignificant terms, the smaller model can be written as follows:

$$Y_{ij} = \beta_{0P} + \beta_{1P}week + \beta_{3P}ind.age + \beta_{0L} + \beta_{1L}week + \beta_{3L}ind.age$$
$$+ \beta_{0H} + \beta_{1H}week + \beta_{3H}ind.age + b_{0i} + b_{1i} + e_{ij}$$

Before going ahead with this smaller model, we should check whether it is significant or not. For that purpose, we will compare this smaller model with our original model.

In general, to compare two nested models, we can use the likelihood ratio test.

To test this, at level α, whether the reduced model is sufficient, we compare the LRT value to the critical value $\chi^2_{df}(\alpha)$ for an α -level test, where df = p1 - p2.

We conclude that the reduced model is sufficient if LRT < $\chi^2_{df}(\alpha)$ and p-value > α

By comparing our original full model with the proposed smaller model, the LRT value and corresponding p-value are as follows:

L.full	L.reduced	LRT	df	p.value
-1511.545	-1515.2	7.309226	15	0.948494

We can also test the same using anova.lme() function. The obtained results are as follows:

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
fit.in.ev	1	28	3079.09	3198.52	-1511.55			
fit.in.ev.s	2	13	3056.40	3111.85	-1515.20	1 Vs 2	7.309226	0.9485

As the p-value from both results is greater than 0.05, we conclude that the terms involving covariate gender and also the interaction terms of the week and ind.age are not needed in the model. The smaller model is sufficient. The AIC/BIC values for this new model are also smaller than the original model.

2.3.3 Comparison of mean trends across three treatment groups

In order to check whether a mean trend is same across three treatment groups or not, we will do a comparison of the mean trends for the following combinations of treatments. Based on the results of the comparison we will make conclusions.

a) Mean trends of Placebo and Succimer- low group
b) Mean trends of Placebo and Succimer- high group
c) Mean trends of Succimer- low and Succimer- high group

Each group has three terms in the equation of mean trajectory i.e., intercept, week and ind.age.

We will compare each of these terms in the mean trend of one group with the corresponding term in the mean trend of the other group using anova.lme() function. If all three terms have a p-value greater than 0.05, we can say that all three terms and hence the mean trends of both groups are equal.

a) <u>Mean trends of Placebo and Succimer- low group</u>

The p-values obtained for comparison of three terms in the mean trend of these groups are as follows:

Terms	p-value
Intercept	0.6192
Week	0.0375
ind.age	0.3856

The p-value of the week is less than 0.05. Hence, we fail to accept the null hypothesis saying terms are equal.
Therefore, Placebo and succimer-low groups have different mean trends.

b) Mean trends of Placebo and Succimer- high group

The p-values obtained for comparison of three terms in the mean trend of these groups are as follows:

Terms	p-value
Intercept	0.9313
Week	0.009
ind.age	0.9918

The p-value of week is less than 0.05. Hence, we fail to accept the null hypothesis saying terms are equal.
Therefore, Placebo and succimer-high groups have different mean trends.

c) Mean trends of Succimer- low and Succimer- high group

The p-values obtained for comparison of three terms in the mean trend of these groups are as follows:

Terms	p-value
Intercept	0.6885
Week	0.5915
ind.age	0.3896

The p-values of all three terms are greater than 0.05. Hence, we accept the null hypothesis saying terms are equal.
Therefore, succimer-low and succimer-high groups have the same mean trends.

From the results of the above three comparisons, we conclude that mean trends are not the same across three treatment groups. But succimer-low and succimer-high groups may have a similar mean trend.

2.3.4 Mean Trends of blood lead level for specific individuals

We have obtained the coefficient estimates of the smaller model. The mean trends of blood lead level can be obtained by multiplying these coefficient estimates by appropriate covariates values.

a) For the placebo treatment group

The mean trend equation for the placebo group (C1=1, C2=0, C3=0) can be written as

Y_{ij} = C1+C1:week+ C1:ind.age

The values of fixed coefficients obtained from the output of model fit are coefficient estimates for corresponding terms.

The equation for a mean trajectory with those estimates becomes,

Y_{ij} =25.077 +(-0.320)*week + (4.40)*ind.age

For age< 24, ind.age =0, therefore equation becomes

Y_{ij} =25.077 – 0.320*week

For age> 24, ind.age =1, therefore equation becomes

Y_{ij} = 25.077– 0.320*week + 4.40

b) For succimer- low treatment group

The mean trend equation for the succimer-low group (C1=0, C2=1, C3=0) can be written as

Y_{ij} =C2+C2:week+ C2:ind.age

The values of fixed coefficients obtained from the output of model fit are coefficient estimates for corresponding terms.

The equation for a mean trajectory with those estimates becomes,

Y_{ij} =24.637 +(-0.8769)*week + (5.504)*ind.age

For age< 24, ind.age =0, therefore equation becomes

Y_{ij} =24.637 – 0.8769*week

For age> 24, ind.age =1, therefore equation becomes

Y_{ij} = 24.637– 0.8769*week + 5.504

c) For succimer- high treatment group

The mean trend equation for the succimer-high group (C0=0, C2=0, C3=1) can be written as

Y_{ij} =C3+C3:week+ C3:ind.age

The values of fixed coefficients obtained from the output of model fit are coefficient estimates for corresponding terms.

The equation for a mean trajectory with those estimates becomes,

Y_{ij} =25.0 +(-1.0205)*week + (4.413)*ind.age

For age< 24, ind.age =0, therefore equation becomes

Y_{ij} =25.0 – 1.0205*week

For age> 24, ind.age =1, therefore equation becomes

Y_{ij} =25.0 – 1.0205*week + 4.413

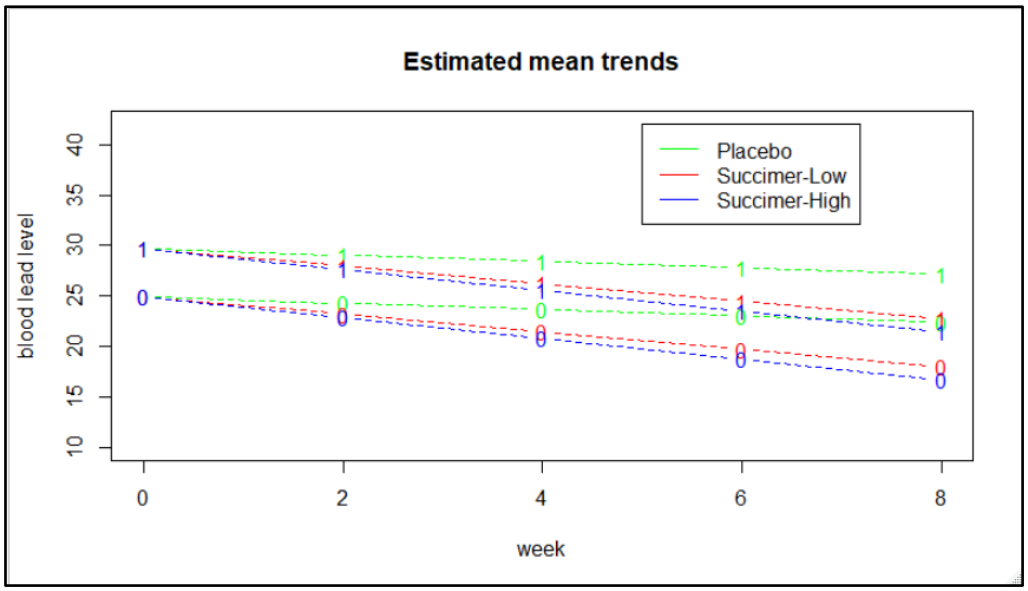
The responses for blood lead levels obtained from these mean trends for all three treatment groups across different weeks are shown in the following tables.

	For Male with age<24				
	Week 0	**Week 2**	**Week 4**	**Week 6**	**Week 8**
Placebo	25.078	24.437	23.797	23.157	22.517
Succimer- Low	24.637	22.883	21.129	19.376	17.622
Succimer- High	25.001	22.960	20.919	18.878	16.837

	For Male with age>24				
	Week 0	**Week 2**	**Week 4**	**Week 6**	**Week 8**
Placebo	29.478	28.838	28.198	27.558	26.917
Succimer- Low	30.142	28.388	26.634	24.880	23.126
Succimer- High	29.415	27.374	25.333	23.291	21.250

As the covariates for gender is not included in the smaller model. The mean trends for females will be the same as the mean trends for males.

The plot for the mean trajectories of across three treatment groups is as follows:



The 0 indicates responses for children having age< 24 months.

The 1 indicates responses for children having age> 24 months.

2.3.5 Model Diagnostics

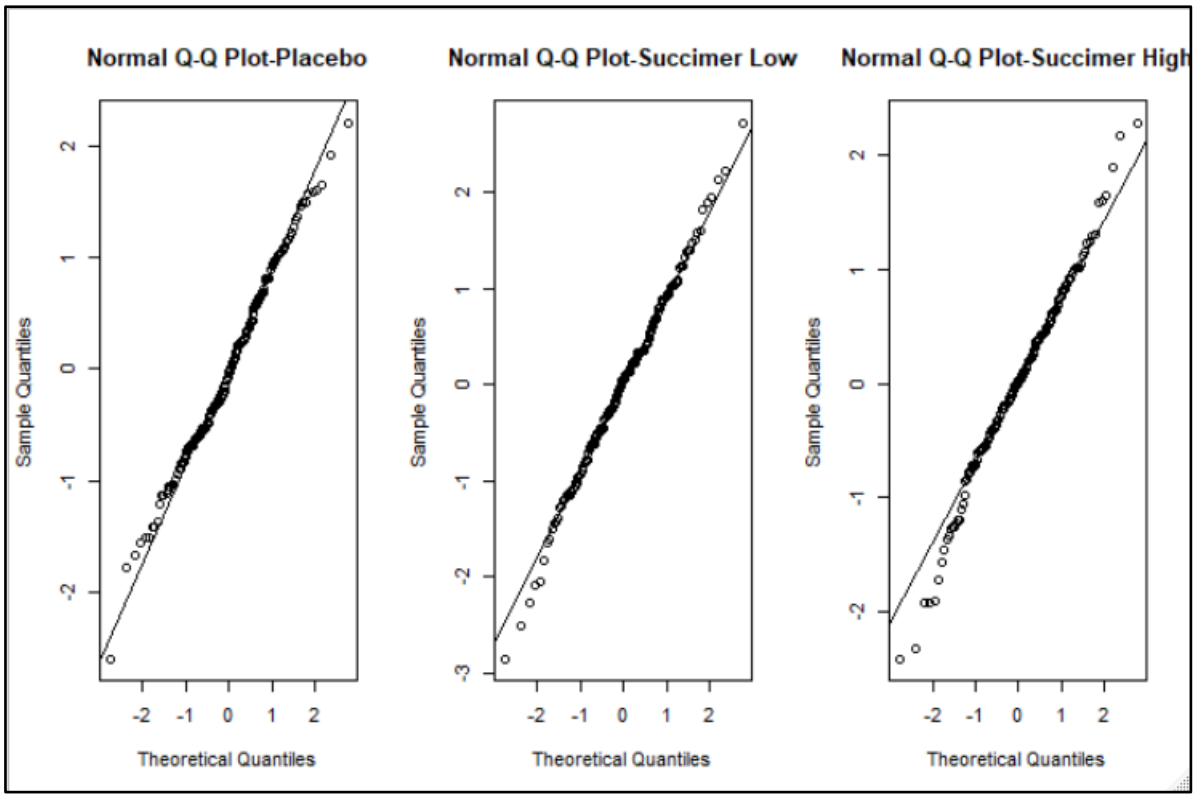
We have considered various assumptions while building a model for our data. We must evaluate these assumptions using model diagnostics. We will asses the validity of our model using various model diagnostics.

The assumptions made for errors are as follows:

1) The errors e_{ij} are normally distributed, and
2) $cov(e_j) = \sigma^2 I$

The error variance is constant and the same for all three treatment groups.

The first assumption i.e., errors are normally distributed can be verified by constructing normal Q-Q plots of the standardized residuals.
The normal Q-Q plots of the standardized residuals for three treatment groups are as follows:



From the above 3 plot, the normality assumption seems to be reasonable. The errors are normally distributed in all three treatment groups.

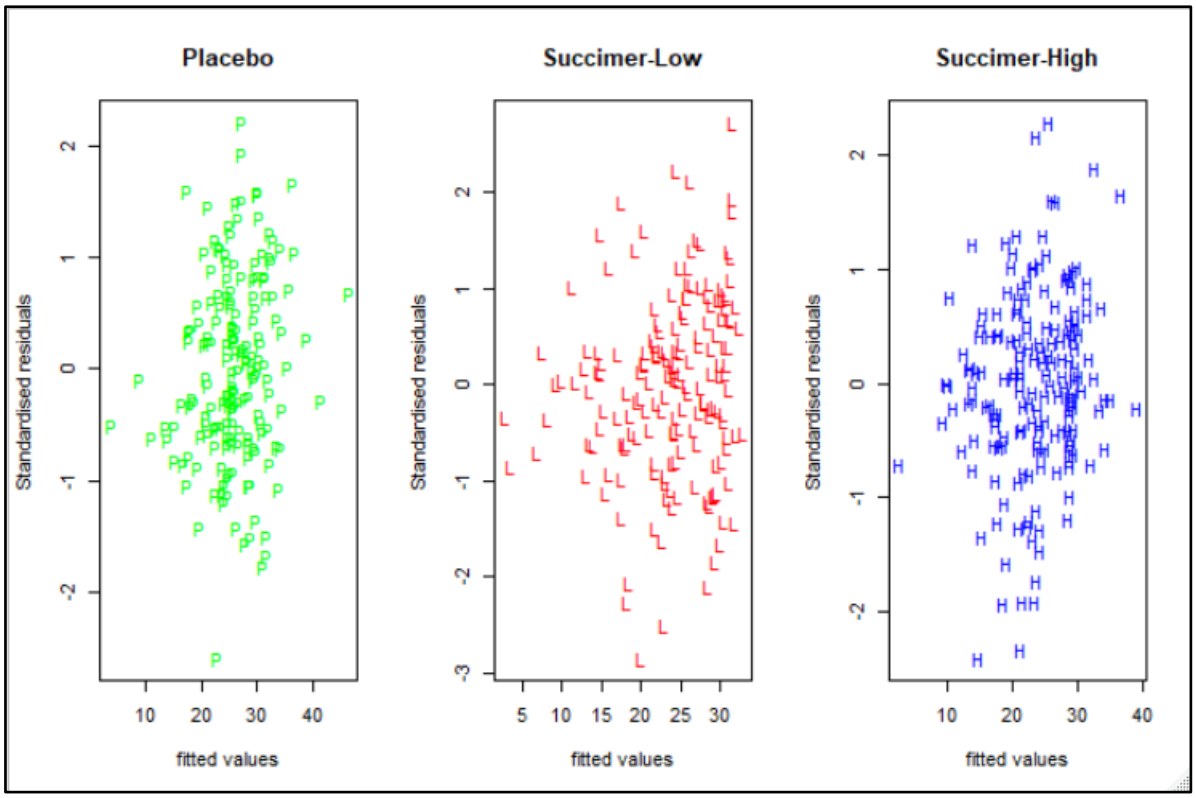
The equal variance assumption can be verified by plotting the Pearson residuals on the y-axis and the fitted values (the subject-level prediction) on the x-axis.

The plot of residuals versus predicted values is useful to check linearity and homoscedasticity assumptions. If the model does not meet the linear model assumption, we would expect to see residuals that are very large (big positive value or big negative value).

To assess the assumption of linearity we want to ensure that the residuals are not too far away from 0.

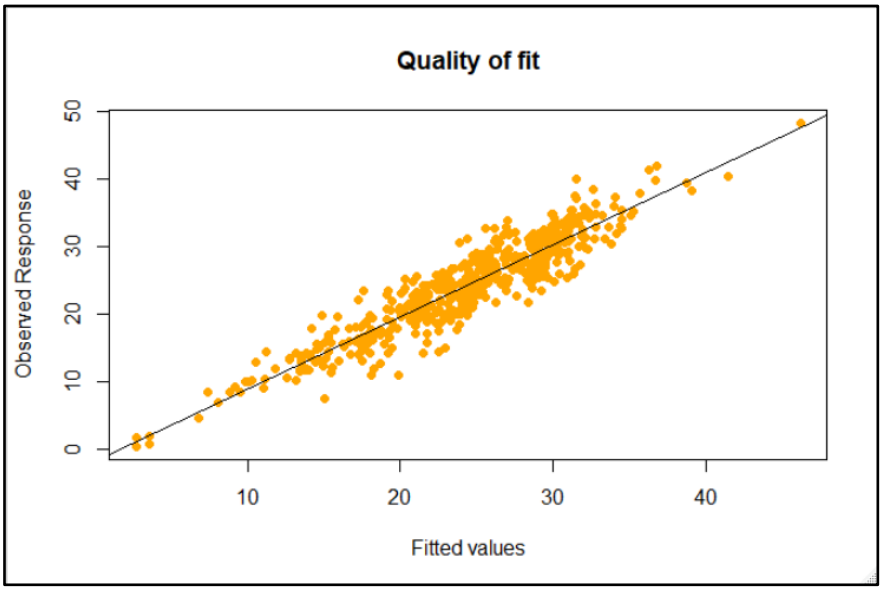
To assess if the homoscedasticity assumption is met we look to make sure that there is no pattern in the residuals and that they are equally spread around the y = 0 line.

The plots for the three treatment groups are as follows:



The residuals for all three treatment groups do not follow any particular shape. They are within +2 and -2 range from 0. Hence, we can say that variance across all groups is equal, the model is linear and there is no heteroscedasticity.

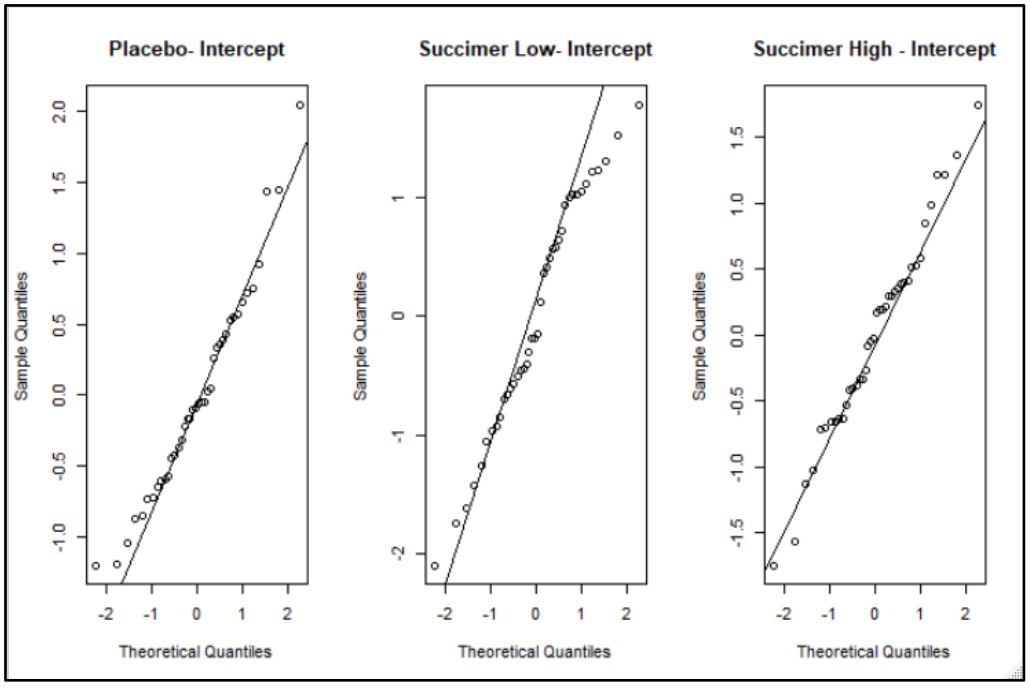
We can also visualize the quality of fit by plotting observed responses vs. subject-level fitted values.

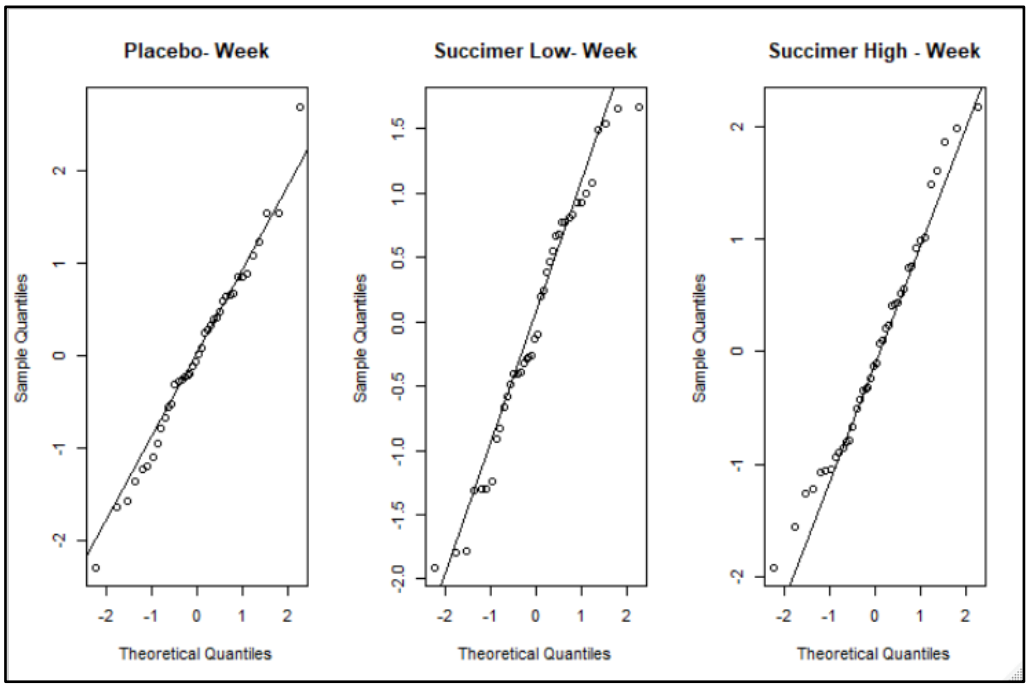


The observed responses are indeed close to the fitted values. Hence we can say that the chosen model is a good fit.

Assumptions regarding random effects:

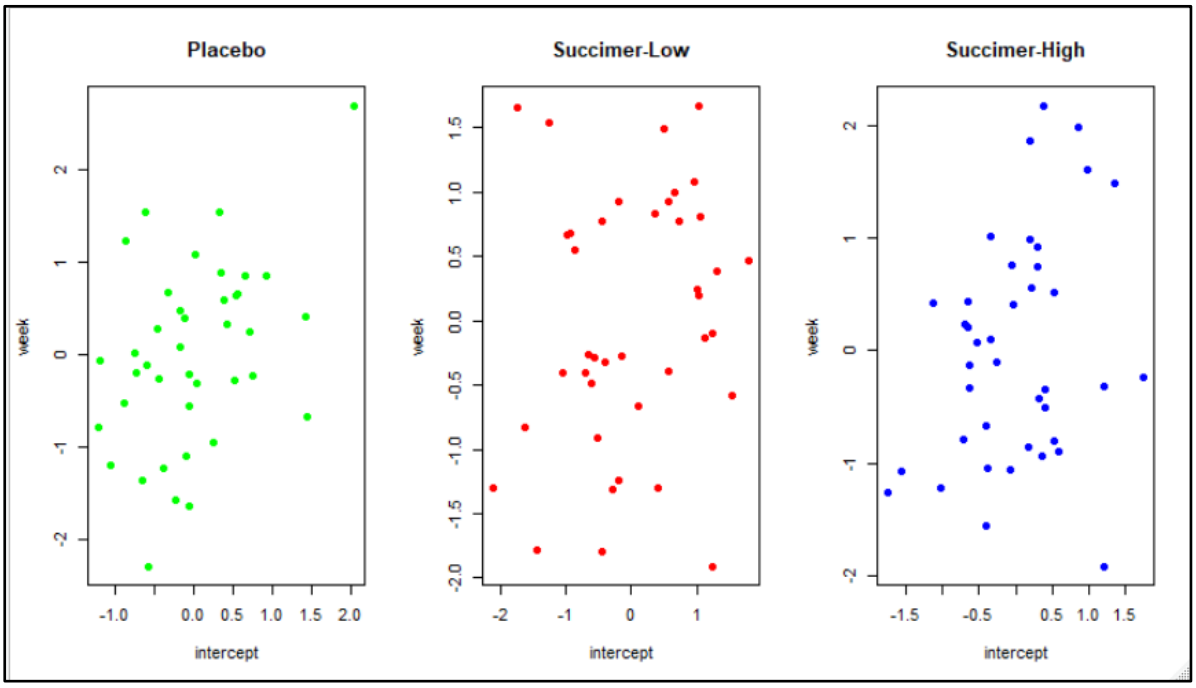
We assume that random effects b_i follows a multivariate normal distribution. We can also construct a normal Q-Q plot of the elements of b_i to evaluate the normality assumption. Q-Q plots of the random effects are shown





From Q-Q plots of both random effects of intercepts and slope of the week, we can say that these random effects follow a multivariate normal distribution.

We can construct scatter plots of the elements of b_i to identify subjects who are different from the rest of the sample. Scatterplots of the random effects for intercept and slope for each of the three groups are shown below.



From the plots, we say that there may be one child receiving placebo treatment which is different from the rest of the sample (top right point). There may not be any outliers in the succimer- low and succimer- high treatment groups.

3. Further reduced model

From the profile plot of the previous smaller model in section 2.3.4, we can see that the intercept for the three treatments group is almost the same. Hence, we can combine the intercept term for all three groups and propose a new smaller model.

The model can be written as follows :

$$Y_{ij} = \beta_0 + \beta_{1P} week + \beta_{3P} ind.age + \beta_{1L} week + \beta_{3L} ind.age + \beta_{1H} week + \beta_{3H} ind.age + b_{0i} + b_{1i} + e_{ij}$$

Also, from section 2.3.3, we see that the p- value of intercept and ind.age across all the comparison is greater than 0.05. Hence, we can combine intercept and ind.age for all three treatment groups into one intercept and one ind.age term.

The smaller model becomes :

$$Y_{ij} = \beta_0 + \beta_{1P} week + \beta_{3P} ind.age + \beta_{1L} week + \beta_3 ind.age + b_{0i} + b_{1i} + e_{ij}$$

3.1 Significance of smaller model

By comparing our original full model with the proposed smaller model, the LRT value and corresponding p-value are as follows:

L.full	L.reduced	LRT	df	p.value
-1511.545	-1515.737	8.383494	19	0.9823787

We can also test the same using anova.lme() function. The obtained results are as follows:

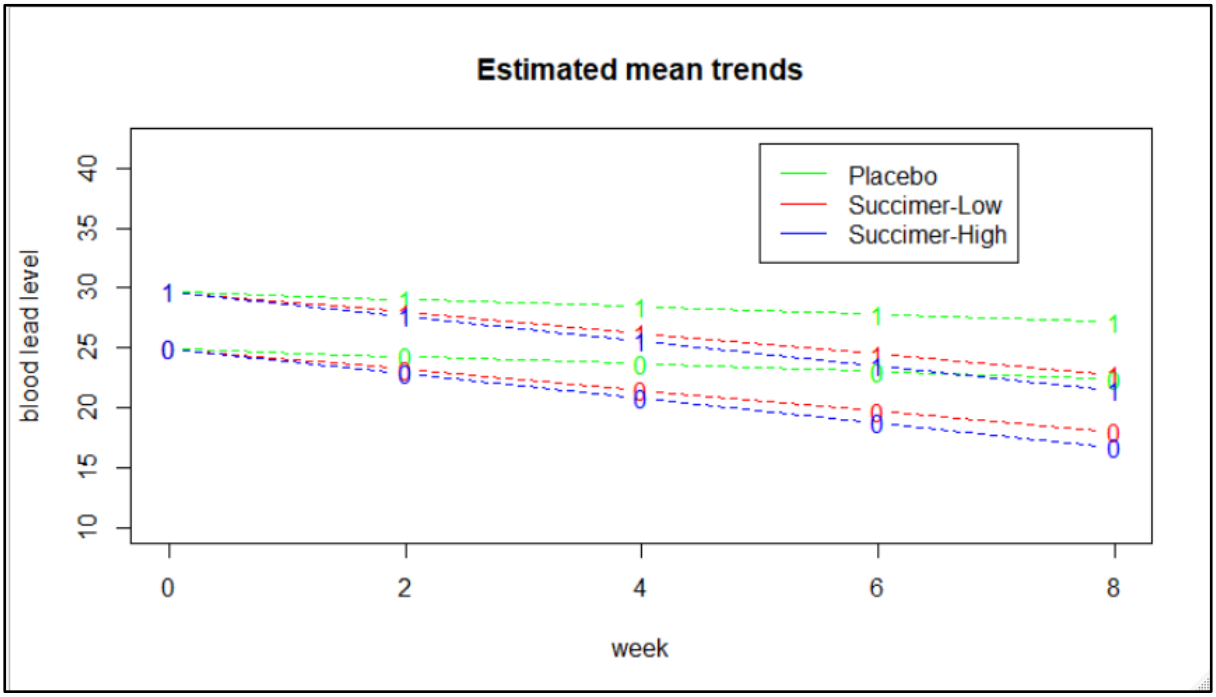
	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
fit.in.ev	1	28	3079.09	3198.52	-1511.545			
fit.in.ev.s	2	9	3049.474	3087.861	-1515.737	1 Vs 2	8.3834	0.9824

As the p-value from both results is greater than 0.05, we conclude that the terms involving covariate gender and also the interaction terms of the week and ind.age are not needed in the model. This smaller model is sufficient. The AIC/BIC values for this new model are also smaller than the original model.

The equation of mean trajectory for this model can be written as :

Y_{ij} = 24.911 + 4.763*ind.age -0.319*C1*week – 0.867*C2*week – 1.0309*C3*week

The plot for the mean trajectories of males across three treatment groups using this new smaller model is as follows:



I figured out that the original smaller model with 9 terms can be further reduced at the last minute. Therefore, with this section, I am trying to show that there is a better-reduced model. But my entire analysis as seen in previous sections was done with a smaller model having 9 terms.

4.Conclusion

The focus of the study was to asses the effect of chelation treatment, succimer, on reducing blood lead levels. Also, it was required to study the association of covariates like age and gender with the blood lead levels. The findings from the analysis of data are as follows:

- Out of three treatment groups, the lowest reduction of blood lead levels was found in children receiving doses of placebo. This is obvious as a placebo is an inactive agent with no lead-lowering properties.
- The mean trends for succimer with low doses and succimer with high doses are almost similar.
- The reduction in blood lead levels for both succimer-low and succimer-high treatments is significant over 8 weeks. Hence, we can conclude that the chelation treatment, succimer is effective in reducing the blood lead levels.
- The blood lead level is associated with the age of the children. The children with lower age (ind.age =0) have lower blood lead levels than children with higher age(ind.age =1)
- The gender of children and its interaction with other covariates has no association with the blood lead level.

5.Appendix

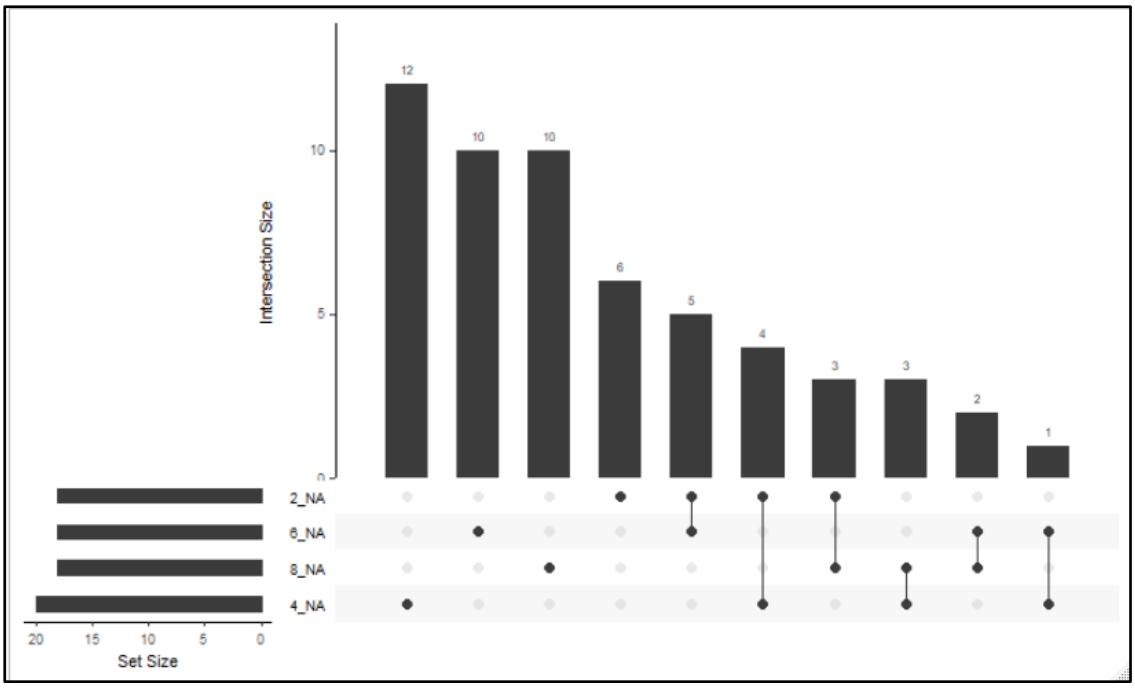
Exploratory data analysis

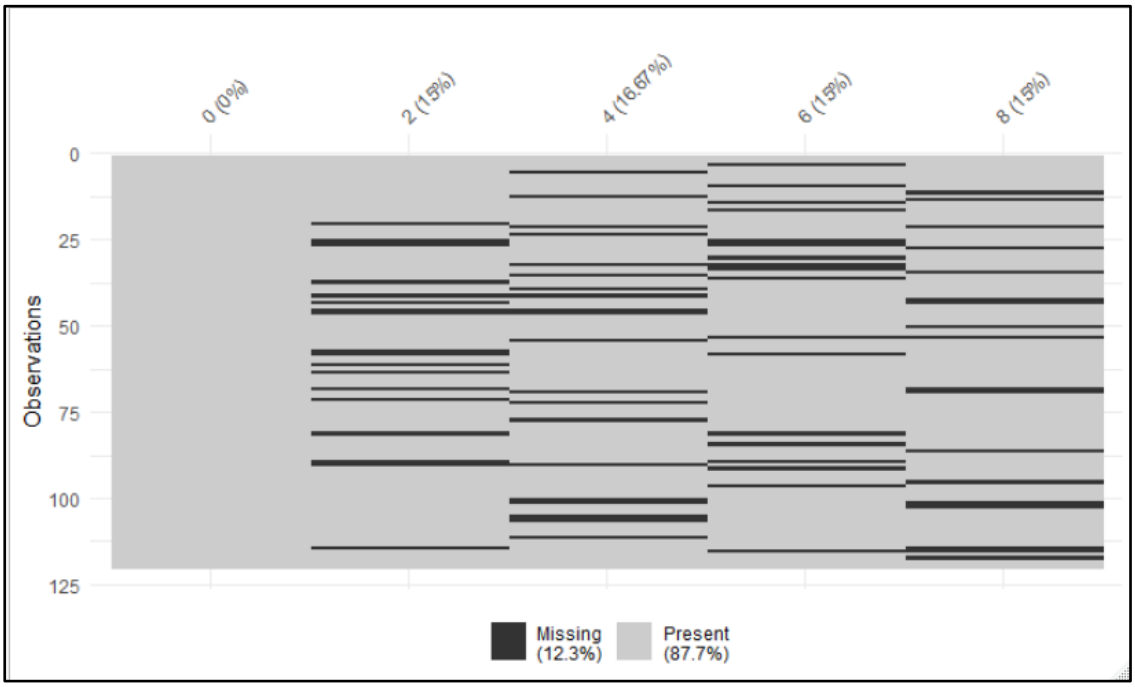
```
#################### Exploratory Data Analysis ##################

###### Reading the Data ######

lead <- read.table("C:/Users/Dell/Dropbox/NCSU/ST 537/Project/lead.full.txt", header = F)
colnames(lead) = c("id", "ind.age", "sex", "week", "blood", "trt")
head(lead)

Y= lead$blood
week = lead$week
ind.age = lead$ind.age
sex= lead$sex

### Indicator variable for groups
C1= as.numeric(lead$trt==1)
C2= as.numeric(lead$trt==2)
C3= as.numeric(lead$trt==3)
lead=cbind(lead,C1,C2,C3)


### Groups
lead1 = lead[C1 == 1,1:7]
lead2 = lead[C2 == 1,1:7]
lead3 = lead[C3 == 1,1:7]

### Converting data to wide format
library(tidyr)
lead_wide<- spread(lead, week, blood)
head(lead_wide)

```

```
> head(lead_wide)
  id ind.age sex trt C1 C2 C3    0    2    4    6    8
1  1       0   1   1  1  0  0 31.8 31.6 39.9 40.5 48.3
2  2       0   0   1  1  0  0 24.5 28.1 30.6 34.8 35.3
3  3       0   1   1  1  0  0 23.9 27.0 22.6   NA 27.7
4  4       0   1   1  1  0  0 26.7 26.3 24.4 26.7 25.5
5  5       0   0   1  1  0  0 25.3 21.5   NA 22.5 26.9
6  6       0   0   1  1  0  0 23.1 25.0 24.3 17.9 17.0
```

```

######################## Missing Values ############################
###################### Number of missing values####################

sum(is.na(lead_wide))

################ Number of missing values across each week and their pattern ##############
library(naniar)

gg_miss_upset(lead_wide[,8:12])

vis_miss(lead_wide[,8:12])

######## Number of children with same number of  missing observations ####################

lead_wide$MissCount = rowSums(is.na(lead_wide))
head(lead_wide)
M=matrix(data=c(0,0,1,0,2,0,3,0,4,0,5,0), nrow=6, byrow = TRUE)
for (i in 0:5){
 M[i+1,2]= nrow(subset(lead_wide,MissCount == i))
  }
colnames(M)=c('No of missing values','No of Entries')
M

```

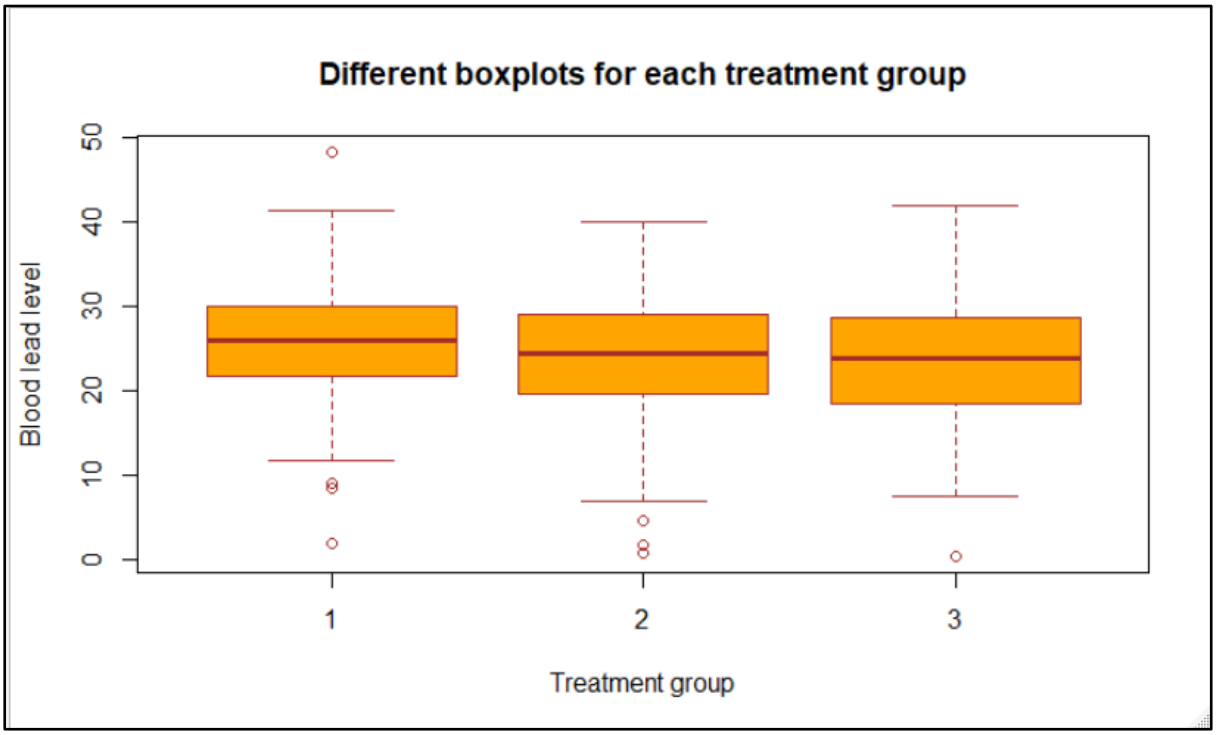
```
> sum(is.na(lead_wide))
[1] 74
>
```






```
> M
     No of missing values No of Entries
[1,]                    0            64
[2,]                    1            38
[3,]                    2            18
[4,]                    3             0
[5,]                    4             0
[6,]                    5             0
>
```

```r
################## Box Plots ###############

boxplot(blood~trt,
        data=lead,
        main="Different boxplots for each treatment group",
        xlab="Treatment group",
        ylab="Blood lead level",
        col="orange",
        border="brown"
)

```




```r

############################## Group 1 Profile plots ##############################################

## Group 1, age<24 months and Male
dat1=lead[(C1 == 1 & ind.age == 0 & sex == 1),]
dat1_wide<- spread(dat1, week, blood)
dat1_wide
y1=colMeans(dat1_wide[,8:12],na.rm = TRUE)


## Group 1, age<24 months and Female
dat2=lead[(C1 == 1 & ind.age == 0 & sex == 0),]
dat2_wide<- spread(dat2, week, blood)
dat2_wide
y2=colMeans(dat2_wide[,8:12],na.rm = TRUE)

### Plotting

xdata= c(0,2,4,6,8)
plot(xdata, y1, type="b", col="blue", pch="M", lty=1, ylim=c(20,30),
     xlab = "week", ylab = "lead", main = 'Estimated mean trends for \n
     Group 1 (Placebo) & age<24 months')

lines(xdata, y2,type = "b", lty = 2, col = "red", pch="F")
legend(6,30,legend=c('Male','Female'),
       col=c("blue", "red"), lty=c(1,1))

#########################################################################################################
```

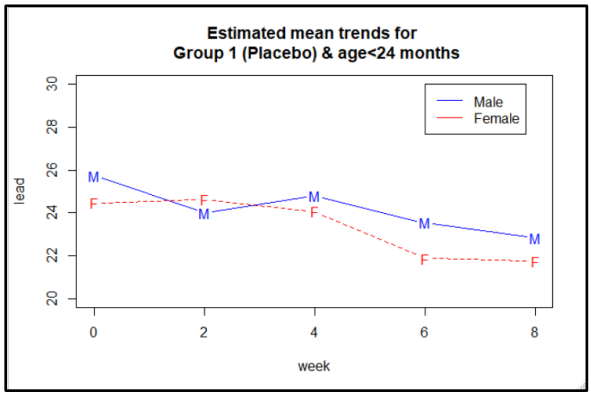
```r
## Group 1, age>24 months and Male
dat3=lead[(C1 == 1 & ind.age == 1 & sex == 1),]
dat3_wide<- spread(dat3, week, blood)
dat3_wide
y3=colMeans(dat3_wide[,8:12],na.rm = TRUE)


## Group 1, age>24 months and Female
dat4=lead[(C1 == 1 & ind.age == 1 & sex == 0),]
dat4_wide<- spread(dat4, week, blood)
dat4_wide
y4=colMeans(dat4_wide[,8:12],na.rm = TRUE)

### Plotting

xdata= c(0,2,4,6,8)
plot(xdata, y3, type="b", col="blue", pch="M", lty=1, ylim=c(25,32),
     xlab = "week", ylab = "lead", main = 'Estimated mean trends for
     \n Group 1(Placebo) & age>24 months')

lines(xdata, y4,type = "b", lty = 2, col = "red", pch="F")
legend(6,31.5,legend=c('Male','Female'),
       col=c("blue", "red"), lty=c(1,1))

```

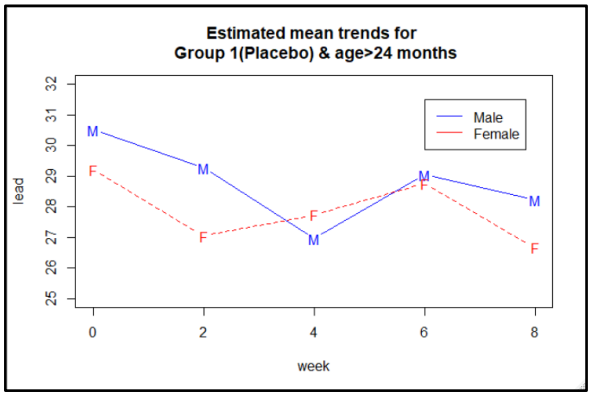
```r
##################################### Group 2 Profile plots#####################################

## Group 2, age<24 months and Male
dat5=lead[(C2 == 1 & ind.age == 0 & sex == 1),]
dat5_wide<- spread(dat5, week, blood)
dat5_wide
y5=colMeans(dat5_wide[,8:12],na.rm = TRUE)


## Group 2, age<24 months and Female
dat6=lead[(C2 == 1 & ind.age == 0 & sex == 0),]
dat6_wide<- spread(dat6, week, blood)
dat6_wide
y6=colMeans(dat6_wide[,8:12],na.rm = TRUE)

### Plotting

xdata= c(0,2,4,6,8)
plot(xdata, y5, type="b", col="blue", pch="M", lty=1, ylim=c(15,26),
     xlab = "week", ylab = "lead", main = 'Estimated mean trends for
     \n Group 2(Succimer-Low) & age<24 months')

lines(xdata, y6,type = "b", lty = 2, col = "red", pch="F")
legend(6,26,legend=c('Male','Female'),
       col=c("blue", "red"), lty=c(1,1))

##############################################################################################

```

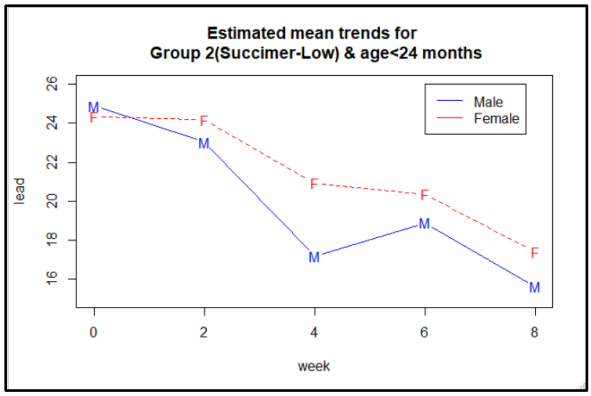
```r
##############################################################################################

## Group 2, age>24 months and Male
dat7=lead[(C2 == 1 & ind.age == 1 & sex == 1),]
dat7_wide<- spread(dat7, week, blood)
dat7_wide
y7=colMeans(dat7_wide[,8:12],na.rm = TRUE)


## Group 2, age>24 months and Female
dat8=lead[(C2 == 1 & ind.age == 1 & sex == 0),]
dat8_wide<- spread(dat8, week, blood)
dat8_wide
y8=colMeans(dat8_wide[,8:12],na.rm = TRUE)

### Plotting

xdata= c(0,2,4,6,8)
plot(xdata, y7, type="b", col="blue", pch="M", lty=1, ylim=c(20,33),
     xlab = "week", ylab = "lead", main = 'Estimated mean trends for
     \n Group 2(Succimer-Low) & age>24 months')

lines(xdata, y8,type = "b", lty = 2, col = "red", pch="F")
legend(6,32.8,legend=c('Male','Female'),
       col=c("blue", "red"), lty=c(1,1))


```

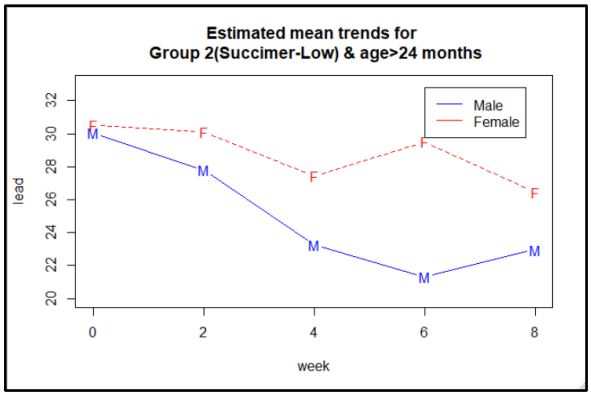
```r
################################### Group 3 Profile plots ######################################

## Group 3, age<24 months and Male
dat9=lead[(C3 == 1 & ind.age == 0 & sex == 1),]
dat9_wide<- spread(dat9, week, blood)
dat9_wide
y9=colMeans(dat9_wide[,8:12],na.rm = TRUE)


## Group 3, age<24 months and Female
dat10=lead[(C3 == 1 & ind.age == 0 & sex == 0),]
dat10_wide<- spread(dat10, week, blood)
dat10_wide
y10=colMeans(dat10_wide[,8:12],na.rm = TRUE)

### Plotting

xdata= c(0,2,4,6,8)
plot(xdata, y9, type="b", col="blue", pch="M", lty=1, ylim=c(15,26),
     xlab = "week", ylab = "lead", main = 'Estimated mean trends for
     \n Group 3(Succimer- High) & age<24 months')

lines(xdata, y10,type = "b", lty = 2, col = "red", pch="F")
legend(6,26,legend=c('Male','Female'),
       col=c("blue", "red"), lty=c(1,1))

#################################################################################################
```

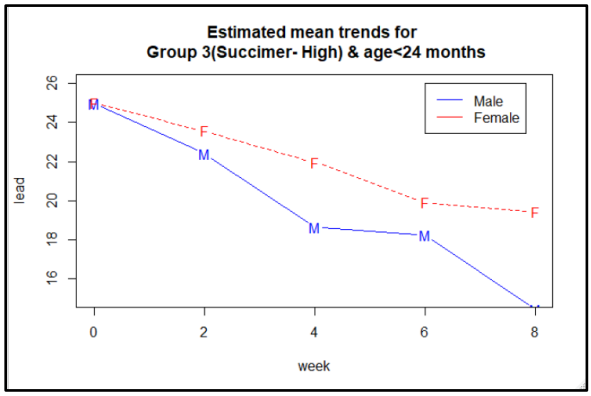
```r
#################################################################################################

## Group 3, age>24 months and Male
dat11=lead[(C3 == 1 & ind.age == 1 & sex == 1),]
dat11_wide<- spread(dat11, week, blood)
dat11_wide
y11=colMeans(dat11_wide[,8:12],na.rm = TRUE)


## Group 3, age>24 months and Female
dat12=lead[(C3 == 1 & ind.age == 1 & sex == 0),]
dat12_wide<- spread(dat12, week, blood)
dat12_wide
y12=colMeans(dat12_wide[,8:12],na.rm = TRUE)

### Plotting

xdata= c(0,2,4,6,8)
plot(xdata, y11, type="b", col="blue", pch="M", lty=1, ylim=c(20,30),
     xlab = "week", ylab = "lead", main = 'Estimated mean trends for
     \n Group 3(Succimer-High) & age>24 months')

lines(xdata, y12,type = "b", lty = 2, col = "red", pch="F")
legend(6,30,legend=c('Male','Female'),
       col=c("blue", "red"), lty=c(1,1))
```

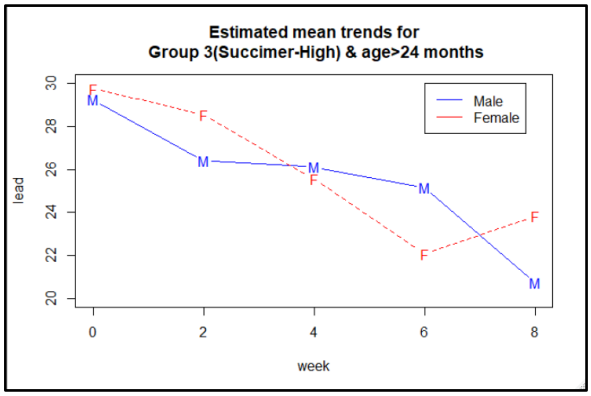




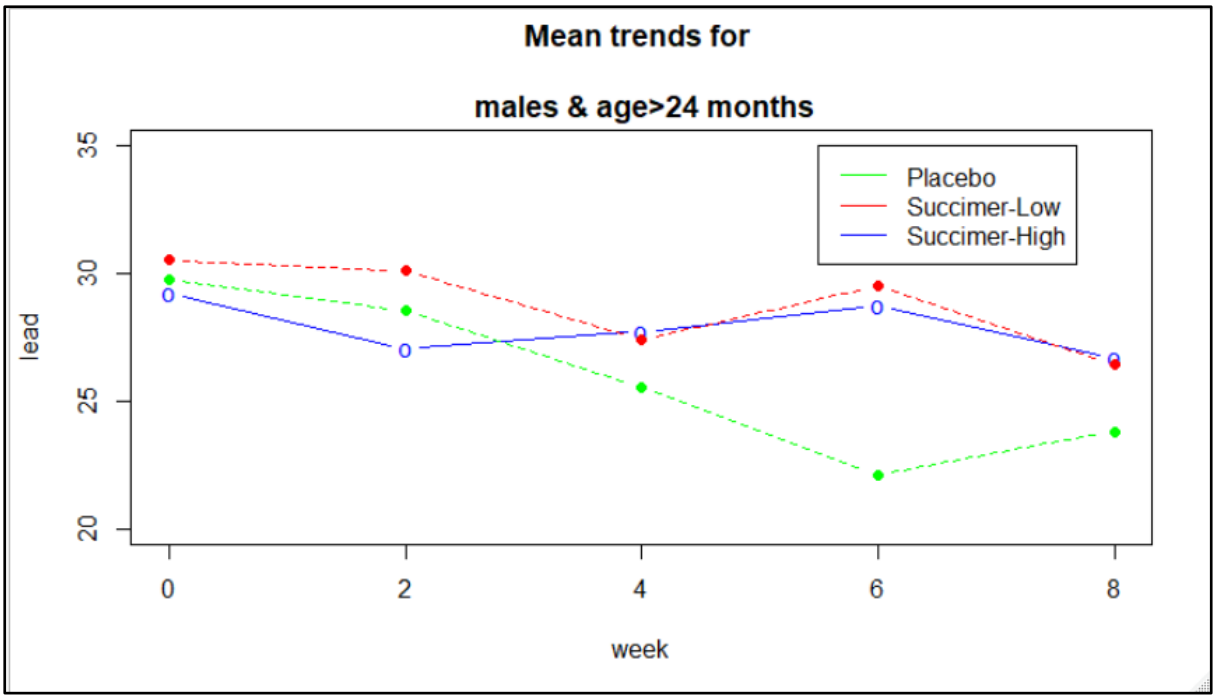






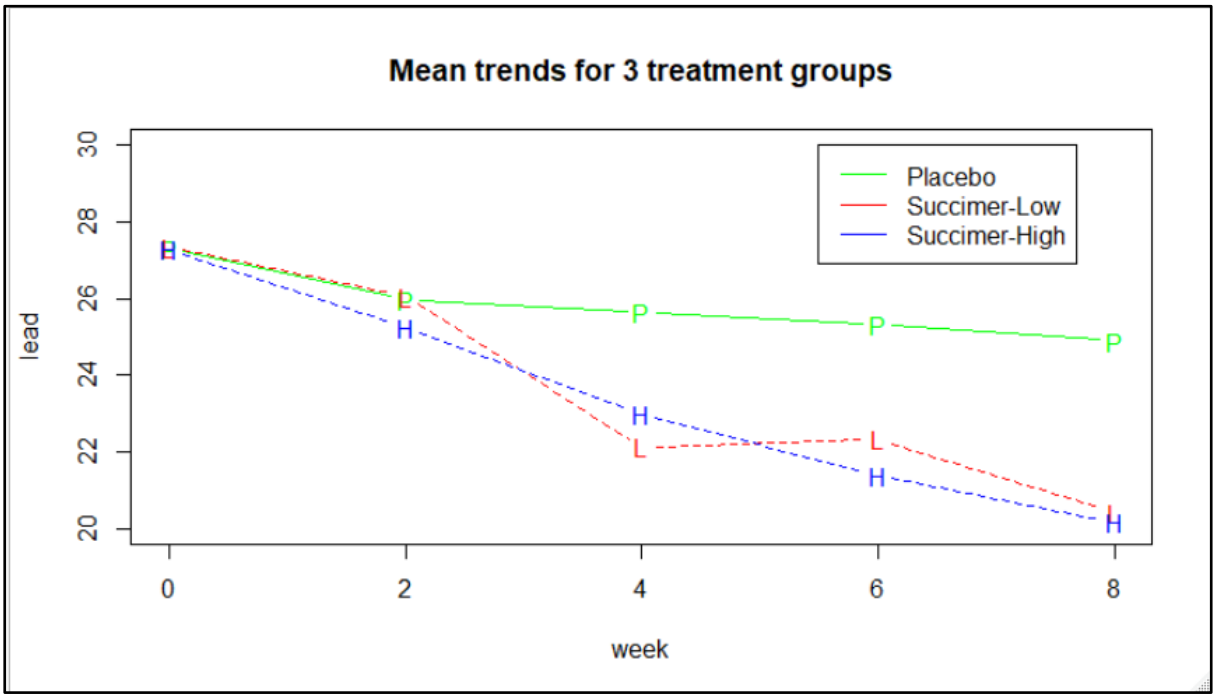



```
################################## Profile plot for males with age > 24 ####################

plot(xdata, y4, type="b", col="blue", pch="o", lty=1, ylim=c(20,35),
     xlab = "week", ylab = "lead", main = 'Mean trends for
     \n males & age>24 months')

lines(xdata, y8,type = "b", lty = 2, col = "red", pch=19)
lines(xdata, y12,type = "b", lty = 2, col = "green", pch=19)

legend(5.5,35,legend=c('Placebo','Succimer-Low','Succimer-High'),
       col=c("green","red", "blue"), lty=c(1,1))
##############################################################################################
```




```r
### Plot given data

par(mfrow=c(1,3))
matplot(lead1$id,lead1$blood,  pch=16, col="green", xlab="ID",
        ylab="lead content", main = 'Placebo group')
plot(lead2$id,lead2$blood,  pch=16, col="red",xlab="ID",
     ylab="lead content",main = 'Succimer-Low group')
plot(lead3$id,lead3$blood,  pch=16, col="blue",xlab="ID",
     ylab="lead content",main = 'Succimer- High group')

require(ggplot2)

## define base for the graphs and store in object 'p'
p <- ggplot(data = lead1, aes(x = week, y = blood, group = id))
## just plotting points (a.k.a., scatterplot)
p + geom_point()
## simple spaghetti plot
p + geom_line()

interaction.plot(lead1[1:60,]$week,
                 lead1[1:60,]$id, lead1[1:60,]$blood,
                 xlab="week", ylab="lead content", col=c(1:10), legend=F,pch = c(1:9, 0, letters))


interaction.plot(lead2[1:60,]$week,
                 lead2[1:60,]$id, lead2[1:60,]$blood,
                 xlab="week", ylab="lead content", col=c(1:10), legend=F,pch = c(1:9, 0, letters))


interaction.plot(lead3[1:60,]$week,
                 lead3[1:60,]$id, lead3[1:60,]$blood,
                 xlab="week", ylab="lead content", col=c(1:10), legend=F,pch = c(1:9, 0, letters))
```

```r
############################ Comparison of three treatments ##############################


### Converting data to wide format
library(tidyr)
lead_wide<- spread(lead, week, blood)
lead_wide

## Number of rows for each treatment
table(lead_wide$trt)


y.p=colMeans(lead_wide[1:40,8:12],na.rm = TRUE)

y.l=colMeans(lead_wide[41:80,8:12],na.rm = TRUE)

y.h=colMeans(lead_wide[81:120,8:12],na.rm = TRUE)



### Plotting

xdata= c(0,2,4,6,8)
plot(xdata, y.p, type="b", col="green", pch="P", lty=1, ylim=c(20,30),
     xlab = "week", ylab = "lead", main = 'Mean trends for 3 treatment groups')

lines(xdata, y.l,type = "b", lty = 2, col = "red", pch="L")
lines(xdata, y.h,type = "b", lty = 2, col = "blue", pch="H")
legend(5.5,30,legend=c('Placebo','Succimer-Low','Succimer-High'),
       col=c("green","red", "blue"), lty=c(1,1))
```



```r
######################## Final Project ######################

#################### Reading the Data ######################

lead <- read.table("C:/Users/Dell/Dropbox/NCSU/ST 537/Project/lead.full.txt", header = F)
colnames(lead) = c("id", "ind.age", "sex", "week", "blood", "trt")
head(lead)

Y= lead$blood
week = lead$week
ind.age = lead$ind.age
sex= lead$sex


### Indicator variable for groups
C1= as.numeric(lead$trt==1)
C2= as.numeric(lead$trt==2)
C3= as.numeric(lead$trt==3)
lead=cbind(lead,C1,C2,C3)


### Treatment Groups
lead1 = lead[C1 == 1,1:7]
lead2 = lead[C2 == 1,1:7]
lead3 = lead[C3 == 1,1:7]

## Number of rows for each treatment
table(lead$trt)

### Models
meanform= blood  ~ -1 + C1 + C1:week + C1:sex + C1:ind.age + C1:week:ind.age+ C1:sex:week+
   C1:sex:ind.age+ C1:sex:week:ind.age+
  (C2+C2:week + C2:sex + C2:ind.age + C2:week:ind.age+ C2:sex:week+
   C2:sex:ind.age+ C2:sex:week:ind.age)+
  (C3+C3:week + C3:sex + C3:ind.age + C3:week:ind.age+ C3:sex:week+
   C3:sex:ind.age+ C3:sex:week:ind.age)
```

```r
##Independent, where error variance does not change over weeks

library(nlme)
fit.in.ev = lme(fixed= meanform,
           data=lead, random = ~ week|id, method="ML",
           control = lmeControl(opt='optim'))
summary(fit.in.ev)
```

```r
### Independent, where error variance changes over weeks

fit.in.uv = lme(fixed= meanform,
           data=lead, random = ~ week|id, method="ML",
           weights = varIdent(form = ~ 1 | week),
           control = lmeControl(opt='optim'))
summary(fit.in.uv )
```

```r
### AR(1) correlation structure, where error variance does not change over weeks

fit.ar1.ev = lme(fixed= meanform,
           data=lead, random = ~ week|id, method="ML",
           correlation = corAR1(form = ~ week | id),
           control = lmeControl(opt='optim'))
summary(fit.ar1.ev)
```

```r
### AR(1) correlation structure, where error variance changes over weeks

fit.ar1.uv = lme(fixed= meanform,
           data=lead, random = ~ week|id, method="ML",
           weights = varIdent(form = ~ 1 | week),
           correlation = corAR1(form = ~ week | id),
           control = lmeControl(opt='optim'))
summary(fit.ar1.uv)
```

```r
### Unstructured, where error variance does not change over weeks

lead$timefact= as.numeric( factor(lead$week, labels = 1:5) )

fit.un.ev = lme(fixed= meanform,
           data=lead, random = ~ week|id, method="ML",
           correlation = corSymm(form= ~ timefact | id ),
           control = lmeControl(opt='optim'))
summary(fit.un.ev)

```

```r
### Unstructured, where error variance changes over weeks

fit.un.uv =lme(fixed= meanform,
           data=lead, random = ~ week|id, method="ML",
           weights = varIdent(form = ~ 1 | timefact),
           correlation = corSymm(form= ~ timefact | id ),
           control = lmeControl(opt='optim'))
summary(fit.un.uv)

```

```r
############## aic and bic ###############

aic = AIC(fit.in.ev,fit.in.uv,fit.ar1.ev,fit.ar1.uv,fit.un.ev, fit.un.uv )
bic = BIC(fit.in.ev,fit.in.uv,fit.ar1.ev,fit.ar1.uv,fit.un.ev, fit.un.uv )
# Display
cbind(aic, bic$BIC)
```

```
> cbind(aic, bic$BIC)
           df      AIC  bic$BIC
fit.in.ev  28 3079.090 3198.518
fit.in.uv  32 3082.369 3218.859
fit.ar1.ev 29 3081.090 3204.784
fit.ar1.uv 33 3084.369 3225.124
fit.un.ev  38 3088.428 3250.509
fit.un.uv  42 3092.238 3271.381
```

```r
############## loglikelihood values ###############

loglik.1 = logLik(fit.in.ev)
loglik.2 = logLik(fit.in.uv)
loglik.3 = logLik(fit.ar1.ev)
loglik.4 = logLik(fit.ar1.uv)
loglik.5 = logLik(fit.un.ev)
loglik.6 = logLik(fit.un.uv)

cbind(loglik.1,loglik.2,loglik.3,loglik.4,loglik.5,loglik.6)
```

```
> cbind(loglik.1,loglik.2,loglik.3,loglik.4,loglik.5,loglik.6)
      loglik.1  loglik.2  loglik.3  loglik.4  loglik.5  loglik.6
[1,] -1511.545 -1509.185 -1511.545 -1509.185 -1506.214 -1504.119
```

```r
################# Robust covariance matrix and se of betahat ######################

library(clubSandwich)
betahat = fixed.effects(fit.in.ev)

# Robust covariance matrix and se of betahat
V.robust = vcovCR(fit.in.ev, type = "CR0")
se.robust = sqrt(diag(V.robust))

# Standard error
##SE = sqrt( diag(fit.a$varFix) )

# CI limits
df = nrow(lead) - length(betahat)
t.alpha = qt(0.05/2, df = df, lower.tail = F)
lower = betahat - t.alpha*se.robust
upper = betahat + t.alpha*se.robust

# Display the estimates
tab = cbind(betahat, se.robust, lower, upper)
tab

```

```
> tab
                          betahat se.robust      lower       upper
C1                    24.99755729 0.7601338 23.5041218 26.49099274
C2                    24.83088516 1.4717496 21.9393374 27.72243290
C3                    25.15581899 0.9011900 23.3852502 26.92638776
C1:week               -0.40651362 0.2281662 -0.8547919  0.04176466
C1:sex                 0.39282851 1.0591286 -1.6880423  2.47369927
C1:ind.age             3.52338472 1.1175727  1.3276887  5.71908078
week:C2               -0.71462367 0.5856148 -1.8651817  0.43593431
sex:C2                -0.13257379 1.8281595 -3.7243602  3.45921263
ind.age:C2             5.68084725 1.7982150  2.1478927  9.21380177
week:C3               -0.89497311 0.2725791 -1.4305095 -0.35943673
sex:C3                -0.23685383 1.4468737 -3.0795278  2.60582017
ind.age:C3             4.58421977 1.0391093  2.5426809  6.62575865
C1:week:ind.age        0.14890086 0.6083927 -1.0464087  1.34421047
C1:week:sex           -0.04502339 0.5499293 -1.1254700  1.03542322
C1:sex:ind.age         0.82133290 1.4758303 -2.0782322  3.72089803
week:ind.age:C2        0.29477716 0.6236293 -0.9304678  1.52002213
week:sex:C2           -0.54107300 0.6726039 -1.8625385  0.78039250
sex:ind.age:C2        -0.84155436 2.3315837 -5.4224187  3.73931002
week:ind.age:C3       -0.06708469 0.4311649 -0.9141947  0.78002531
week:sex:C3           -0.49204706 0.5893909 -1.6500238  0.66592971
sex:ind.age:C3        -0.61768574 1.7133613 -3.9839282  2.74855674
C1:week:sex:ind.age    0.12931039 0.8364165 -1.5139979  1.77261867
week:sex:ind.age:C2   -0.02116744 0.7769436 -1.5476292  1.50529437
week:sex:ind.age:C3    0.52150608 0.7976130 -1.0455649  2.08857701
>
```

```r
################## Effect of age on blood lead level ######################

# L matrix
La =  matrix(0, 12, 24)
La[1,6]=1
La[2,9]=1
La[3,12]=1
La[4,13]=1
La[5,15]=1
La[6,16]=1
La[7,18]=1
La[8,19]=1
La[9,21]=1
La[10,22]=1
La[11,23]=1
La[12,24]=1
# Estimate and SE
estimate = La %*% betahat
SE = La %*% V.robust %*% t(La)
# confidence limits
df = nrow(lead) - length(betahat)
t.alpha = qt(0.05/2, df = df, lower.tail = F)
lower = estimate - t.alpha*SE
upper = estimate + t.alpha*SE
# results
tab = data.frame(estimate, SE, lower, upper)
round(tab, 4)

```

```r
################### Hypothesis testing ######################

########## wald test #########
cc = nrow(La)
df = nrow(lead) - length(betahat)
# estimate and covariance matrix of L\beta
est = La %*% betahat
varmat = La %*% V.robust %*% t(La)
# Wald test
Wald = c( t(est) %*% solve(varmat) %*% (est) )
p.value = pchisq(q = Wald, df = cc, lower.tail=FALSE)
data.frame(Wald, p.value)

```

```
> data.frame(Wald, p.value)
      Wald     p.value
1 98.07974 1.32269e-15
```

```r
########## t stat ###############

# df
df = nrow(lead) - length(betahat)
# t stats
t.robust = betahat/se.robust
# p-values
p.value = round( 2*pt(q = abs(t.robust), df = df, lower.tail = FALSE), 4 )
# results
tab = data.frame(betahat, se.robust, t.robust, p.value)
tab

```

```
> tab
                          betahat se.robust    t.robust p.value
C1                    24.99755729 0.7601338 32.88573468  0.0000
C2                    24.83088516 1.4717496 16.87167751  0.0000
C3                    25.15581899 0.9011900 27.91400118  0.0000
C1:week               -0.40651362 0.2281662 -1.78165592  0.0754
C1:sex                 0.39282851 1.0591286  0.37089786  0.7109
C1:ind.age             3.52338472 1.1175727  3.15271183  0.0017
week:C2               -0.71462367 0.5856148 -1.22029637  0.2229
sex:C2                -0.13257379 1.8281595 -0.07251763  0.9422
ind.age:C2             5.68084725 1.7982150  3.15915908  0.0017
week:C3               -0.89497311 0.2725791 -3.28335198  0.0011
sex:C3                -0.23685383 1.4468737 -0.16370042  0.8700
ind.age:C3             4.58421977 1.0391093  4.41168206  0.0000
C1:week:ind.age        0.14890086 0.6083927  0.24474466  0.8068
C1:week:sex           -0.04502339 0.5499293 -0.08187123  0.9348
C1:sex:ind.age         0.82133290 1.4758303  0.55652258  0.5781
week:ind.age:C2        0.29477716 0.6236293  0.47268013  0.6366
week:sex:C2           -0.54107300 0.6726039 -0.80444520  0.4215
sex:ind.age:C2        -0.84155436 2.3315837 -0.36093681  0.7183
week:ind.age:C3       -0.06708469 0.4311649 -0.15558941  0.8764
week:sex:C3           -0.49204706 0.5893909 -0.83483995  0.4042
sex:ind.age:C3        -0.61768574 1.7133613 -0.36051108  0.7186
C1:week:sex:ind.age    0.12931039 0.8364165  0.15460047  0.8772
week:sex:ind.age:C2   -0.02116744 0.7769436 -0.02724450  0.9783
week:sex:ind.age:C3    0.52150608 0.7976130  0.65383348  0.5135
```

```
################## Effect of Gender on blood lead level ######################


# L matrix
Lg=  matrix(0, 12, 24)
Lg[1,5]=1
Lg[2,8]=1
Lg[3,11]=1
Lg[4,14]=1
Lg[5,15]=1
Lg[6,17]=1
Lg[7,18]=1
Lg[8,20]=1
Lg[9,21]=1
Lg[10,22]=1
Lg[11,23]=1
Lg[12,24]=1
Lg
# Estimate and SE
estimate.g = Lg %*% betahat
SE.g = Lg %*% V.robust %*% t(Lg)
# confidence limits
df = nrow(lead) - length(betahat)
t.alpha = qt(0.05/2, df = df, lower.tail = F)
lower.g = estimate.g - t.alpha*SE.g
upper.g = estimate.g + t.alpha*SE.g
# results
tab = data.frame(estimate.g, SE.g, lower.g, upper.g)
round(tab, 4)
```

```
################### Hypothesis testing ######################

########## wald test #########
cc = nrow(Lg)
df = nrow(lead) - length(betahat)
# estimate and covariance matrix of L\beta
est.g = Lg %*% betahat
varmat.g = Lg %*% V.robust %*% t(Lg)
# Wald test
Wald.g = c( t(est.g) %*% solve(varmat.g) %*% (est.g) )
p.value.g = pchisq(q = Wald.g, df = cc, lower.tail=FALSE)
data.frame(Wald.g, p.value.g)
```

```
> data.frame(Wald.g, p.value.g)
   Wald.g p.value.g
1 8.23386 0.7665984
```

```r
########## t stat ###############

# df
df = nrow(lead) - length(betahat)
# t stats
t.robust = betahat/se.robust
# p-values
p.value = round( 2*pt(q = abs(t.robust), df = df, lower.tail = FALSE), 4 )
# results
tab = data.frame(betahat, se.robust, t.robust, p.value)
tab

```

```
> tab
                       betahat se.robust    t.robust p.value
C1                  24.99755729 0.7601338 32.88573468  0.0000
C2                  24.83088516 1.4717496 16.87167751  0.0000
C3                  25.15581899 0.9011900 27.91400118  0.0000
C1:week             -0.40651362 0.2281662 -1.78165592  0.0754
C1:sex               0.39282851 1.0591286  0.37089786  0.7109
C1:ind.age           3.52338472 1.1175727  3.15271183  0.0017
week:C2             -0.71462367 0.5856148 -1.22029637  0.2229
sex:C2              -0.13257379 1.8281595 -0.07251763  0.9422
ind.age:C2           5.68084725 1.7982150  3.15915908  0.0017
week:C3             -0.89497311 0.2725791 -3.28335198  0.0011
sex:C3              -0.23685383 1.4468737 -0.16370042  0.8700
ind.age:C3           4.58421977 1.0391093  4.41168206  0.0000
C1:week:ind.age      0.14890086 0.6083927  0.24474466  0.8068
C1:week:sex         -0.04502339 0.5499293 -0.08187123  0.9348
C1:sex:ind.age       0.82133290 1.4758303  0.55652258  0.5781
week:ind.age:C2      0.29477716 0.6236293  0.47268013  0.6366
week:sex:C2         -0.54107300 0.6726039 -0.80444520  0.4215
sex:ind.age:C2      -0.84155436 2.3315837 -0.36093681  0.7183
week:ind.age:C3     -0.06708469 0.4311649 -0.15558941  0.8764
week:sex:C3         -0.49204706 0.5893909 -0.83483995  0.4042
sex:ind.age:C3      -0.61768574 1.7133613 -0.36051108  0.7186
C1:week:sex:ind.age  0.12931039 0.8364165  0.15460047  0.8772
week:sex:ind.age:C2 -0.02116744 0.7769436 -0.02724450  0.9783
week:sex:ind.age:C3  0.52150608 0.7976130  0.65383348  0.5135
```

```r
############################## Smaller model ################################

### Models
meanform.s= blood ~ -1 + C1 + C1:week  +  C1:ind.age+
  (C2+C2:week  +  C2:ind.age)+
  (C3+C3:week +   C3:ind.age)


##Independent, where error variance does not change over weeks

library(nlme)
fit.in.ev.s = lme(fixed= meanform.s,
                  data=lead, random = ~ week|id, method="ML",
                  control = lmeControl(opt='optim'))
summary(fit.in.ev.s)

```

```
> summary(fit.in.ev.s)
Linear mixed-effects model fit by maximum likelihood
 Data: lead
       AIC      BIC    logLik
  3056.399 3111.848 -1515.2

Random effects:
 Formula: ~week | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev   Corr
(Intercept) 1.529901 (Intr)
week        1.054789 -0.02
Residual    3.117382

Fixed effects: list(meanform.s)
               Value Std.Error  DF  t-value p-value
C1         25.077706 0.6089649 114 41.18087  0.0000
C2         24.637117 0.6408389 114 38.44510  0.0000
C3         25.001330 0.6408362 114 39.01360  0.0000
C1:week    -0.320138 0.1882362 404 -1.70072  0.0898
C1:ind.age  4.400701 0.8918409 114  4.93440  0.0000
week:C2    -0.876930 0.1889665 404 -4.64067  0.0000
ind.age:C2  5.504658 0.9007358 114  6.11129  0.0000
week:C3    -1.020587 0.1892492 404 -5.39282  0.0000
ind.age:C3  4.413616 0.8856843 114  4.98328  0.0000
```

```
 Correlation:
           C1     C2     C3     C1:wek C1:nd. wek:C2 in.:C2 wek:C3
C2          0.000
C3          0.000  0.000
C1:week    -0.237  0.000  0.000
C1:ind.age -0.644  0.000  0.000 -0.003
week:C2     0.000 -0.241  0.000  0.000  0.000
ind.age:C2  0.000 -0.671  0.000  0.000  0.000  0.004
week:C3     0.000  0.000 -0.235  0.000  0.000  0.000  0.000
ind.age:C3  0.000  0.000 -0.685  0.000  0.000  0.000  0.000  0.004

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-2.8516552 -0.5421721  0.0139481  0.5577853  2.7273523

Number of Observations: 526
Number of Groups: 120
```

```
341 ###################### Significance of Smaller model ########################
342 # Full model
343 fit.in.ev = lme(fixed= meanform,
344                  data=lead, random = ~ week|id, method="ML",
345                  control = lmeControl(opt='optim'))
346 p1 = 28
347 # Reduced model
348 fit.in.ev.s = lme(fixed= meanform.s,
349                    data=lead, random = ~ week|id, method="ML",
350                    control = lmeControl(opt='optim'))
351 p2 = 13
352 # log-likelihoods
353 loglik.full = logLik(fit.in.ev)
354 loglik.red = logLik(fit.in.ev.s)
355 # LRT and p-value
356 df = p1 - p2
357 LRT = 2*(loglik.full - loglik.red)
358 p.value = pchisq(LRT, df = df, lower.tail = FALSE)
359 # results
360 data.frame(L.full = loglik.full, L.reduced = loglik.red,
361            LRT = LRT, df = df, p.value = p.value)
362
```

```
> data.frame(L.full = loglik.full, L.reduced = loglik.red,
+            LRT = LRT, df = df, p.value = p.value)
     L.full L.reduced      LRT df  p.value
1 -1511.545   -1515.2 7.309226 15 0.948494
>
```

```r
############################ Using anova.lme()#######################

anova.lme(fit.in.ev, fit.in.ev.s)

```

```
> anova.lme(fit.in.ev, fit.in.ev.s)
           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fit.in.ev      1 28 3079.090 3198.518 -1511.545
fit.in.ev.s    2 13 3056.399 3111.848 -1515.200 1 vs 2 7.309226  0.9485
>
```

```r
################# Robust covariance matrix and se of betahat ######################

library(clubSandwich)
betahat.s = fixed.effects(fit.in.ev.s)

# Robust covariance matrix and se of betahat
V.robust.s = vcovCR(fit.in.ev.s, type = "CR0")
se.robust.s = sqrt(diag(V.robust.s))

# Standard error
##SE = sqrt( diag(fit.a$varFix) )

# CI limits
df = nrow(lead) - length(betahat.s)
t.alpha = qt(0.05/2, df = df, lower.tail = F)
lower = betahat.s - t.alpha*se.robust.s
upper = betahat.s + t.alpha*se.robust.s

# Display the estimates
tab = cbind(betahat.s, se.robust.s, lower, upper)
tab
```

```
> tab
             betahat.s se.robust.s      lower       upper
C1          25.0777063   0.5580520 23.9813779 26.17403462
C2          24.6371171   0.8267587 23.0128975 26.26133667
C3          25.0013297   0.6691892 23.6866653 26.31599412
C1:week     -0.3201378   0.1781847 -0.6701929  0.02991723
C1:ind.age   4.4007011   0.7460310  2.9350762  5.86632603
week:C2     -0.8769303   0.1944011 -1.2588436 -0.49501697
ind.age:C2   5.5046581   1.0603367  3.4215597  7.58775655
week:C3     -1.0205875   0.1897772 -1.3934168 -0.64775819
ind.age:C3   4.4136164   0.7971071  2.8476492  5.97958350
>
```

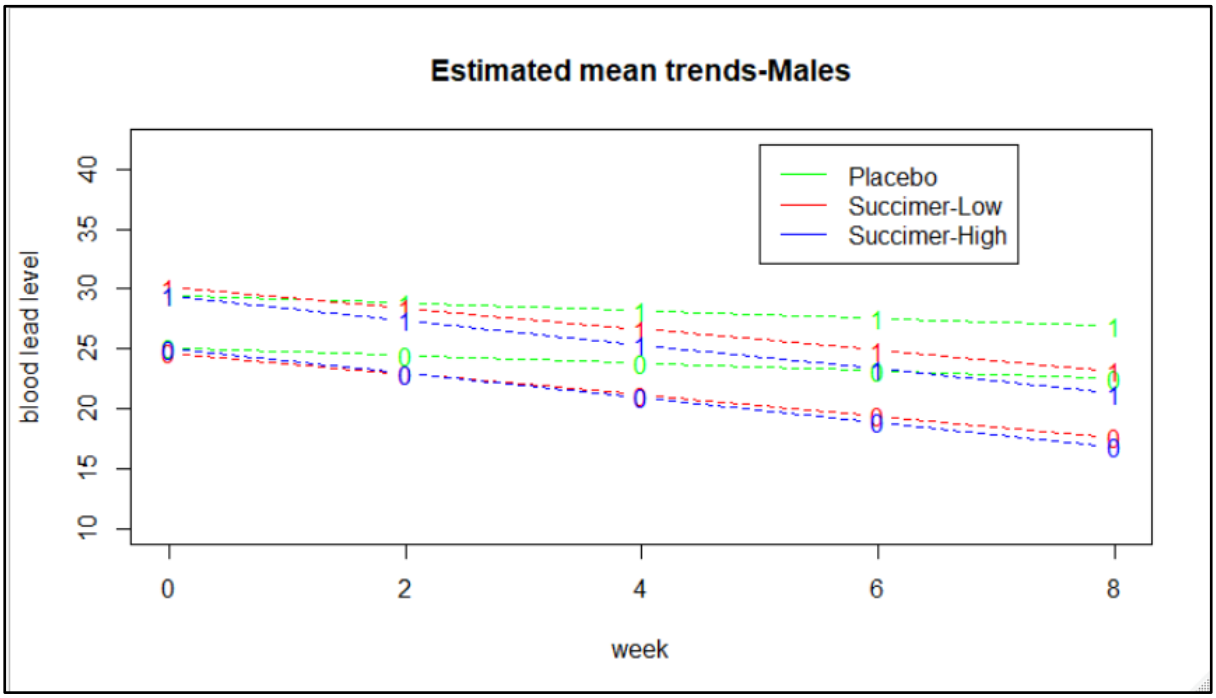
```r
###################### Comparison of mean trends ######################

### Placebo and Succimer- Low
anova.lme(fit.in.ev.s, L = c(1,-1,0,0,0,0,0,0,0), adjustSigma = TRUE)
anova.lme(fit.in.ev.s, L = c(0,0,0,1,0,-1,0,0,0), adjustSigma = TRUE)
anova.lme(fit.in.ev.s, L = c(0, 0,0,0,1,0,-1,0,0), adjustSigma = TRUE)

### Placebo and Succimer- High
anova.lme(fit.in.ev.s, L = c(1, 0,-1,0,0,0,0,0,0), adjustSigma = TRUE)
anova.lme(fit.in.ev.s, L = c(0,0,0,1,0,0,0,-1,0), adjustSigma = TRUE)
anova.lme(fit.in.ev.s, L = c(0, 0,0,0,1,0,0,0,-1), adjustSigma = TRUE)

### Succimer- Low and Succimer -High
anova.lme(fit.in.ev.s, L = c(0, 1,-1,0,0,0,0,0,0), adjustSigma = TRUE)
anova.lme(fit.in.ev.s, L = c(0,0,0,0,0,1,0,-1,0), adjustSigma = TRUE)
anova.lme(fit.in.ev.s, L = c(0, 0,0,0,0,0,1,0,-1), adjustSigma = TRUE)

```

```
> ### Placebo and Succimer- Low
> anova.lme(fit.in.ev.s, L = c(1,-1,0,0,0,0,0,0,0), adjustSigma = TRUE)
F-test for linear combination(s)
C1 C2
 1 -1
  numDF denDF   F-value p-value
1     1   114 0.2483886  0.6192
> anova.lme(fit.in.ev.s, L = c(0,0,0,1,0,-1,0,0,0), adjustSigma = TRUE)
F-test for linear combination(s)
C1:week week:C2
      1      -1
  numDF denDF  F-value p-value
1     1   404 4.357781  0.0375
> anova.lme(fit.in.ev.s, L = c(0, 0,0,0,1,0,-1,0,0), adjustSigma = TRUE)
F-test for linear combination(s)
C1:ind.age ind.age:C2
         1         -1
  numDF denDF   F-value p-value
1     1   114 0.7585218  0.3856
>
```

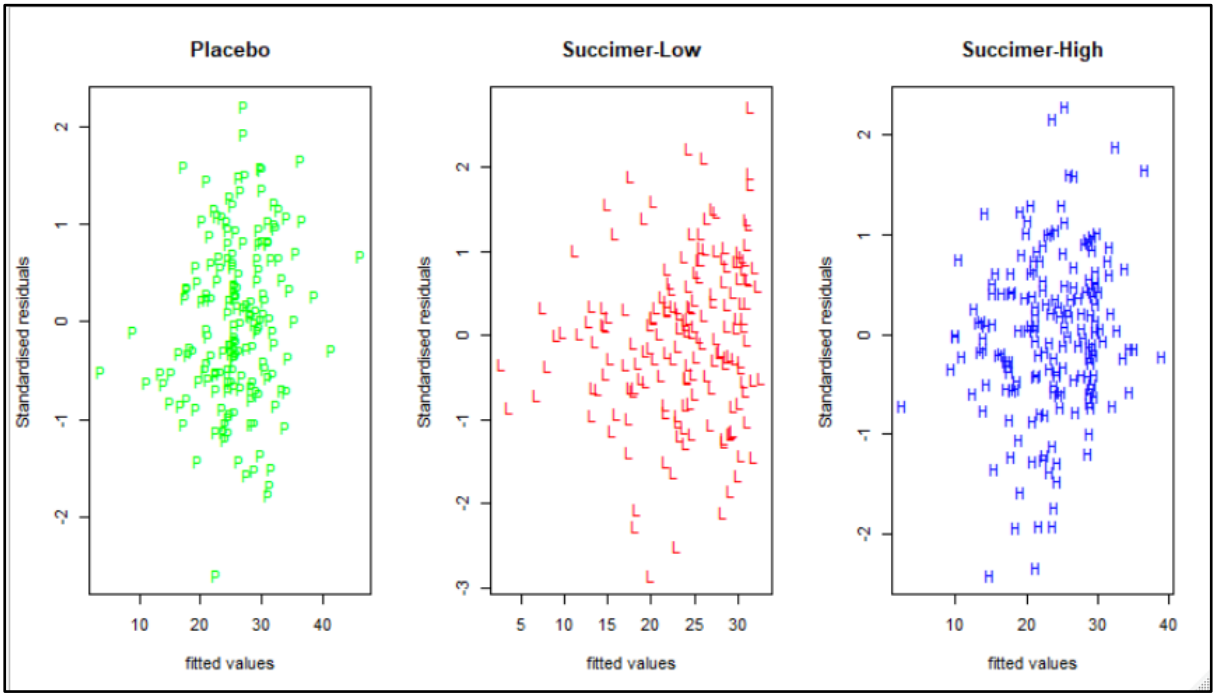
```
> ### Placebo and Succimer- High
> anova.lme(fit.in.ev.s, L = c(1, 0,-1,0,0,0,0,0,0), adjustSigma = TRUE)
F-test for linear combination(s)
C1 C3
 1 -1
  numDF denDF     F-value p-value
1     1   114 0.007464246  0.9313
> anova.lme(fit.in.ev.s, L = c(0,0,0,1,0,0,0,-1,0), adjustSigma = TRUE)
F-test for linear combination(s)
C1:week week:C3
      1      -1
  numDF denDF F-value p-value
1     1   404 6.88621   0.009
> anova.lme(fit.in.ev.s, L = c(0, 0,0,0,1,0,0,0,-1), adjustSigma = TRUE)
F-test for linear combination(s)
C1:ind.age ind.age:C3
         1         -1
  numDF denDF      F-value p-value
1     1   114 0.0001055839  0.9918
>
```

```
> ### Succimer- Low and Succimer -High
> anova.lme(fit.in.ev.s, L = c(0, 1,-1,0,0,0,0,0,0), adjustSigma = TRUE)
F-test for linear combination(s)
C2 C3
 1 -1
  numDF denDF   F-value p-value
1     1   114 0.1615043  0.6885
> anova.lme(fit.in.ev.s, L = c(0,0,0,0,0,1,0,-1,0), adjustSigma = TRUE)
F-test for linear combination(s)
week:C2 week:C3
      1      -1
  numDF denDF   F-value p-value
1     1   404 0.2885396  0.5915
> anova.lme(fit.in.ev.s, L = c(0, 0,0,0,0,0,1,0,-1), adjustSigma = TRUE)
F-test for linear combination(s)
ind.age:C2 ind.age:C3
         1         -1
  numDF denDF   F-value p-value
1     1   114 0.7459585  0.3896
```

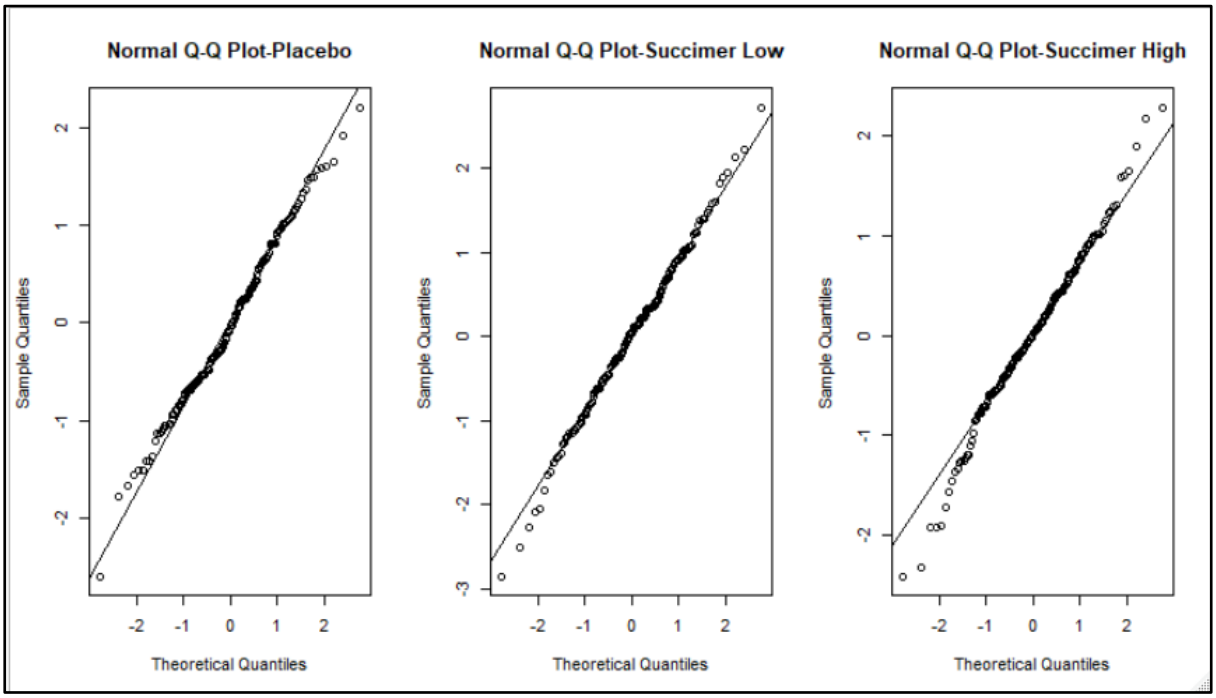
```r
############################ Mean Trends ############################

cf= fit.in.ev.s$coefficients$fixed
response <- function(c1,c2,c3,t,ag){

    u= c1*cf[1]+c1*cf[4]*t+c1*cf[5]*ag+
      c2*cf[2]+c2*cf[6]*t+c2*cf[7]*ag+
      c3*cf[3]+c3*cf[8]*t+c3*cf[9]*ag
 u
}   ### function to get estimated value


response <- Vectorize(response)
wk <- c(0,2,4,6,8)

P.Male.A0 = response(1,0,0,wk,0)
P.Male.A1 = response(1,0,0,wk,1)
SL.Male.A0 = response(0,1,0,wk,0)
SL.Male.A1 = response(0,1,0,wk,1)
SH.Male.A0 = response(0,0,1,wk,0)
SH.Male.A1 = response(0,0,1,wk,1)


matplot(c(0,2,4,6,8), P.Male.A0, type = "b", lty = 2, col = "green", pch="0",
        xlab = "week", ylab = "blood lead level", main = 'Estimated mean trends-Males',
        ylim = c(10,42))
lines(wk, P.Male.A1 , type = "b", lty = 2, col = "green", pch="1",)
lines(wk, SL.Male.A0, type = "b", lty = 2, col = "red", pch="0",)
lines(wk, SL.Male.A1, type = "b", lty = 2, col = "red", pch="1",)
lines(wk, SH.Male.A0, type = "b", lty = 2, col = "blue", pch="0",)
lines(wk, SH.Male.A1, type = "b", lty = 2, col = "blue", pch="1",)

legend(5,42,legend=c('Placebo','Succimer-Low','Succimer-High'),
       col=c("green","red", "blue"), lty=c(1,1))
```



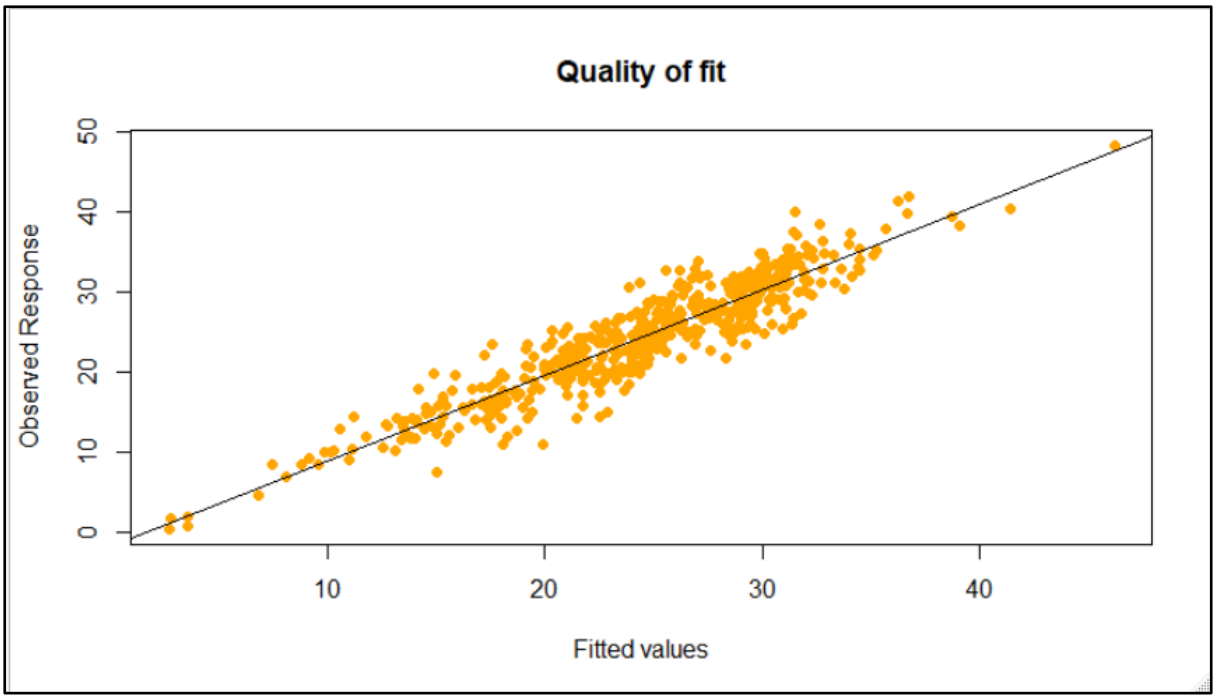
```r
############################## Model Diagnostics ###################################

################### Checking equal variance assumption ################

############ Subject level predictions - Fitted values ################

sub.pred = predict(fit.in.ev.s, level = 1)
P.pred = sub.pred[1:174]
SL.pred = sub.pred[175:349]
SH.pred = sub.pred[350:526]

############# Pearson Residuals ##############

res = resid(fit.in.ev.s, level = 1, type = "pearson")
P.res = res[1:174]
SL.res = res[175:349]
SH.res = res[350:526]

par(mfrow=c(1,3))

matplot(P.pred, P.res, col = "green", pch="P",
        xlab = "fitted values", ylab = "Standardised residuals", main = 'Placebo')
matplot(SL.pred, SL.res, col = "red", pch="L",
        xlab = "fitted values", ylab = "Standardised residuals", main = 'Succimer-Low')
matplot(SH.pred, SH.res, col = "blue", pch="H",
        xlab = "fitted values", ylab = "Standardised residuals", main = 'Succimer-High')

```



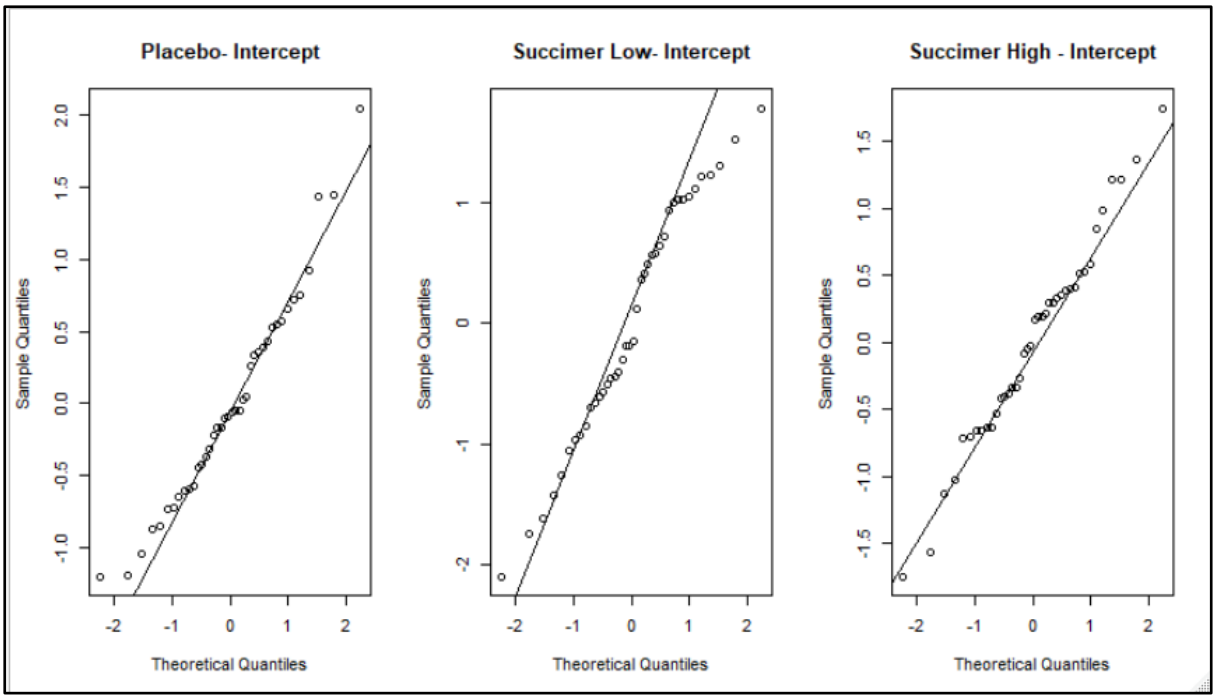
```r
################### Checking assumption of normal errors ################

par(mfrow=c(1,3))

qqnorm(P.res, main = "Normal Q-Q Plot-Placebo",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
       plot.it = TRUE, datax = FALSE)
qqline(P.res)

qqnorm(SL.res, main = "Normal Q-Q Plot-Succimer Low",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
       plot.it = TRUE, datax = FALSE)
qqline(SL.res)

qqnorm(SH.res, main = "Normal Q-Q Plot-Succimer High",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
       plot.it = TRUE, datax = FALSE)
qqline(SH.res)
```



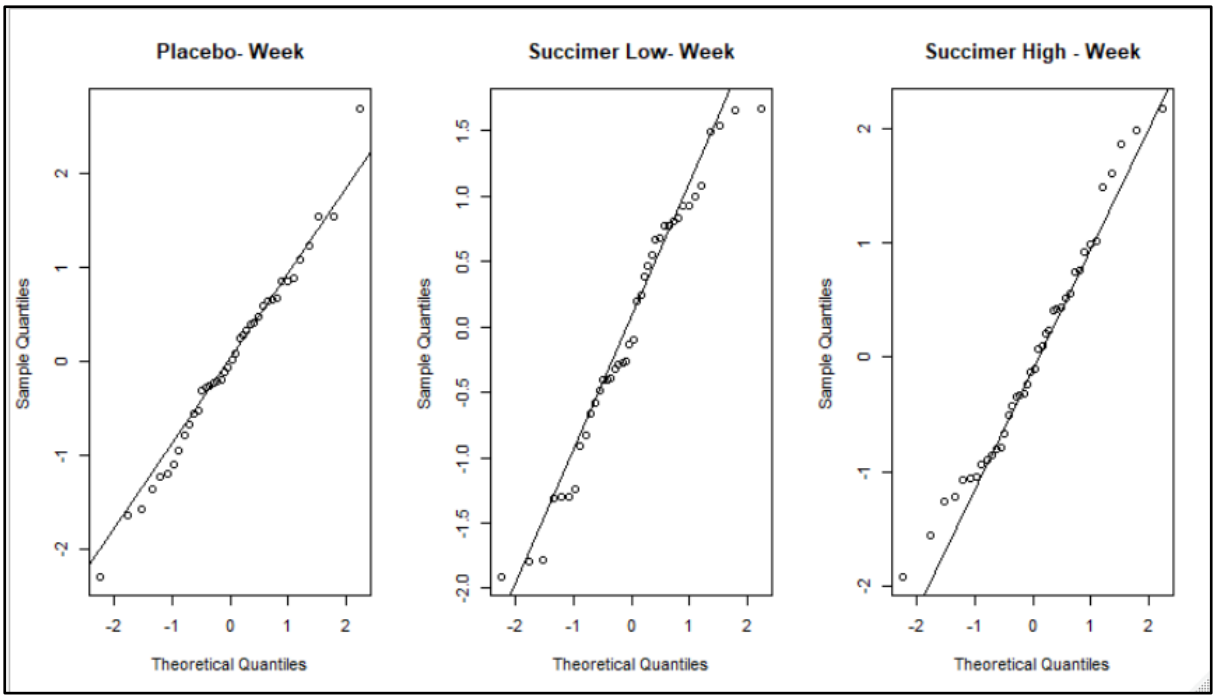

```r
################ Quality of fit ###################


matplot(sub.pred, lead$blood, col = "orange", pch=16,
        xlab = "Fitted values", ylab = "Observed Response", main = 'Quality of fit')
abline(lm(lead$blood~ sub.pred))
```



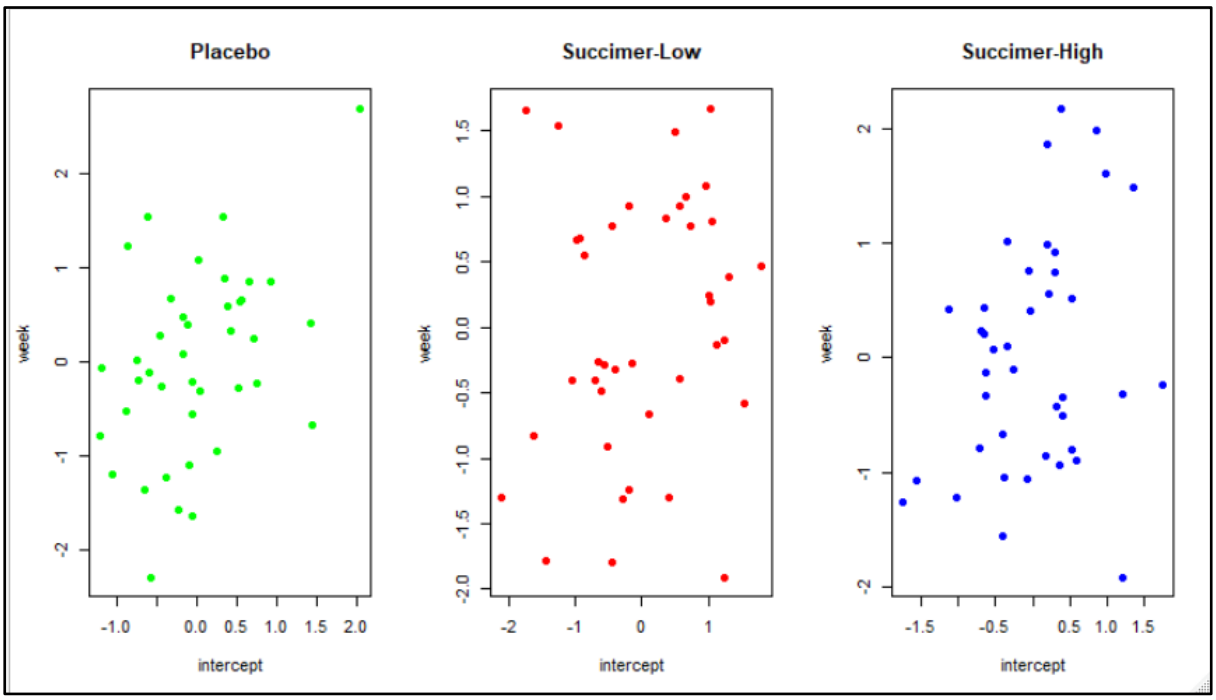

```r
################ Prediction of random effects ##################

b.hat = random.effects(fit.in.ev.s,)
b.hat[1:3, ]

################### Q-Q plots for random effects ################

################### Q-Q plots for intercepts ################

par(mfrow=c(1,3))
qqnorm(b.hat[1:40,1], main = "Placebo- Intercept",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
       plot.it = TRUE, datax = FALSE)
qqline(b.hat[1:40,1])

qqnorm(b.hat[41:80,1], main = "Succimer Low- Intercept",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
       plot.it = TRUE, datax = FALSE)
qqline(b.hat[41:80,1])

qqnorm(b.hat[81:120,1], main = "Succimer High - Intercept",
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
       plot.it = TRUE, datax = FALSE)
qqline(b.hat[81:120,1])

```




```
528 ##################### Q-Q plots for weeks ########################
529
530 par(mfrow=c(1,3))
531
532 qqnorm(b.hat[1:40,2], main = "Placebo- Week",
533        xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
534        plot.it = TRUE, datax = FALSE)
535 qqline(b.hat[1:40,2])
536
537 qqnorm(b.hat[41:80,2], main = "Succimer Low- Week",
538        xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
539        plot.it = TRUE, datax = FALSE)
540 qqline(b.hat[41:80,2])
541
542 qqnorm(b.hat[81:120,2], main = "Succimer High - Week",
543        xlab = "Theoretical Quantiles", ylab = "Sample Quantiles",
544        plot.it = TRUE, datax = FALSE)
545 qqline(b.hat[81:120,2])
546
```




```
547 ############### Scatter plots for random effects ######################
548
549 par(mfrow=c(1,3))
550
551 plot(b.hat[1:40,1],b.hat[1:40,2],main = "Placebo",
552      xlab = "intercept", ylab = "week",col= "green",pch = 19,
553      plot.it = TRUE, datax = FALSE)
554
555 plot(b.hat[41:80,1],b.hat[41:80,2],main = "Succimer-Low",
556      xlab = "intercept", ylab = "week",col= "red",pch=19,
557      plot.it = TRUE, datax = FALSE)
558
559 plot(b.hat[81:120,1],b.hat[81:120,2],main = "Succimer-High",
560      xlab = "intercept", ylab = "week",col= "blue",pch=19,
561      plot.it = TRUE, datax = FALSE)
562
563
```



```r
#################### New Smaller model ##############################


### Models
meanform.s2= blood ~   C1:week  + ind.age+ C2:week  +C3:week


##Independent, where error variance does not change over weeks

library(nlme)
fit.in.ev.s2 = lme(fixed= meanform.s2,
                   data=lead, random = ~ week|id, method="ML",
                   control = lmeControl(opt='optim'))
summary(fit.in.ev.s2)

```

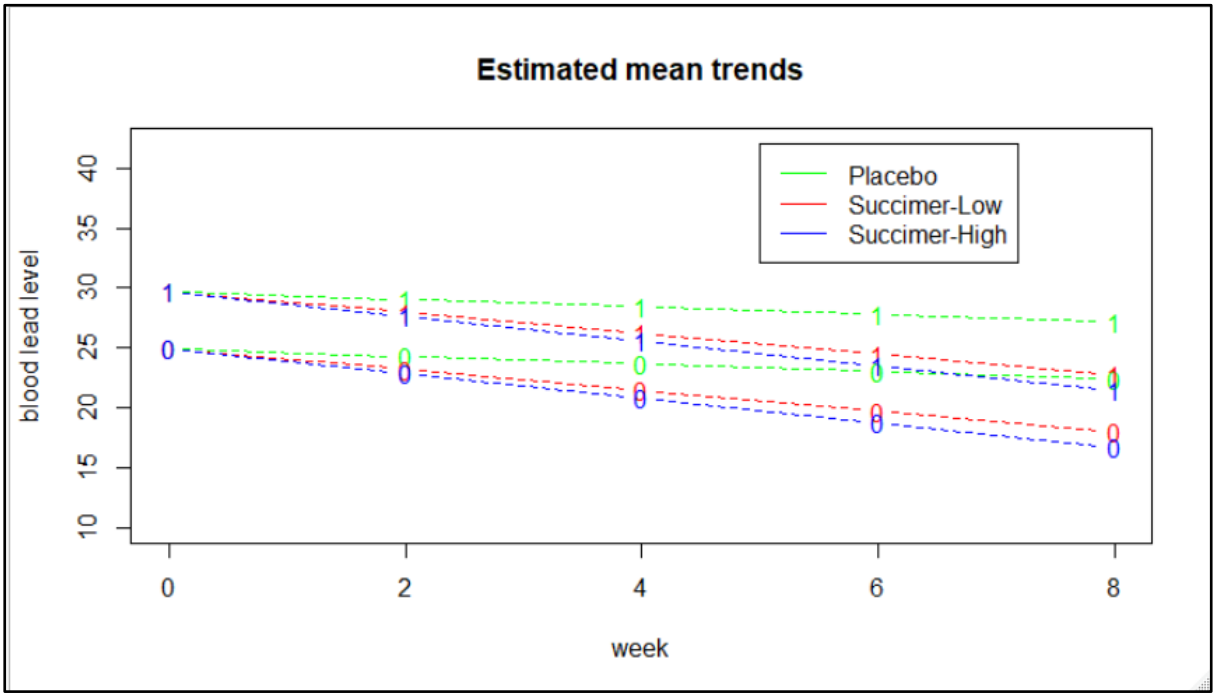
```r
#################### Significance of Smaller model ###################
# Full model
fit.in.ev = lme(fixed= meanform,
                data=lead, random = ~ week|id, method="ML",
                control = lmeControl(opt='optim'))

p1 <- 28
# Reduced model
fit.in.ev.s2 = lme(fixed= meanform.s2,
                   data=lead, random = ~ week|id, method="ML",
                   control = lmeControl(opt='optim'))
p2 <- 9
# log-likelihoods
loglik.full <- logLik(fit.in.ev)
loglik.red <- logLik(fit.in.ev.s2)
# LRT and p-value
df <- p1 - p2
LRT <- 2*(loglik.full - loglik.red)
p.value <- pchisq(LRT, df = df, lower.tail = FALSE)
# results
data.frame(L.full = loglik.full, L.reduced = loglik.red,
           LRT = LRT, df = df, p.value = p.value)

```

```
> data.frame(L.full = loglik.full, L.reduced = loglik.red,
+            LRT = LRT, df = df, p.value = p.value)
     L.full L.reduced      LRT df   p.value
1 -1511.545 -1515.737 8.383494 19 0.9823787
>
```

```r
######### Using anova.lme()
anova.lme(fit.in.ev, fit.in.ev.s2)

```

```
> anova.lme(fit.in.ev, fit.in.ev.s2)
            Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fit.in.ev       1 28 3079.090 3198.518 -1511.545
fit.in.ev.s2    2  9 3049.474 3087.861 -1515.737 1 vs 2 8.383494  0.9824
>
```

```r
############################## Mean Trends ##############################

cf= fit.in.ev.s2$coefficients$fixed
response <- function(c1,c2,c3,t,ag){

  u= cf[1]+ag*cf[2]+c1*cf[3]*t+c2*cf[4]*t+c3*cf[5]*t
  u
}   ### function to get estimated value


response <- Vectorize(response)
wk <- c(0,2,4,6,8)

P.Male.A0 = response(1,0,0,wk,0)
P.Male.A1 = response(1,0,0,wk,1)
SL.Male.A0 = response(0,1,0,wk,0)
SL.Male.A1 = response(0,1,0,wk,1)
SH.Male.A0 = response(0,0,1,wk,0)
SH.Male.A1 = response(0,0,1,wk,1)


matplot(c(0,2,4,6,8), P.Male.A0, type = "b", lty = 2, col = "green", pch="0",
        xlab = "week", ylab = "blood lead level", main = 'Estimated mean trends',
        ylim = c(10,42))
lines(wk, P.Male.A1 , type = "b", lty = 2, col = "green", pch="1",)
lines(wk, SL.Male.A0, type = "b", lty = 2, col = "red", pch="0",)
lines(wk, SL.Male.A1, type = "b", lty = 2, col = "red", pch="1",)
lines(wk, SH.Male.A0, type = "b", lty = 2, col = "blue", pch="0",)
lines(wk, SH.Male.A1, type = "b", lty = 2, col = "blue", pch="1",)

legend(5,42,legend=c('Placebo','Succimer-Low','Succimer-High'),
       col=c("green","red", "blue"), lty=c(1,1))
```



6.References

https://stats.idre.ucla.edu/r/faq/how-can-i-make-spaghetti-plots/

https://stackoverflow.com/questions/2564258/plot-two-graphs-in-same-plot-in-r