Concepts of Multilevel, Longitudinal and Mixed Models - G0B76A

May 24, 2021

Group L

Antoine Chen r0863011
Armen Abagyan r0823936
Juli Dema r0821709
Olga Grytsenko r0360270

Introduction

The data set consists of measurements taken on 401 patients suffering from severe headaches obtained from a longitudinal randomized clinical trial. Out of these patients, 205 are treated with acupuncture and 196 are assigned a placebo treatment. Since the groups are unequal, the design of the study is unbalanced. Patients are followed over a one year period and measurements are taken at three different time points: at baseline (0 months), after 3 months and after 12 months. There are missing measurements, as summarized by Table 1. The dropout rate is similar for both groups, and the patients who dropped out are highly comparable to the completers (Vickers et al, 2004). Thus, we assume that the dropouts are not due to a particular reason and continue the analysis without issues.

	Group = 0 (placebo)		**Group = 1 (acupuncture)**	
Time	**Observations**	**Missing Obs.**	**Observations**	**Missing Obs.**
0	196	0	205	0
3	153	43	173	32
12	140	56	159	44

Table 1: Number of observations and missing observations for each treatment group.

Mean Structure

As shown in Figure 1 (left), the mean of the placebo group is higher at baseline. At the end of the study, this relationship holds but the difference is larger, suggesting the effectiveness of acupuncture as a treatment for severe headaches. Further, the slope for the acupuncture group appears to be steeper. The mean evolution is nonlinear and thus using a model assuming linear evolution may be inappropriate. However, taking a transformation of time ($logtime = ln(time + 1)$) makes the relationship approximately linear as also shown in Figure 1 (right). This transformation will be used in the following analysis and thus proceeding plots are also displayed in this scale.

It is informative to look at the individual profiles shown in Figure 2. There is no obvious difference in the intercepts of both groups, which can be expected as the treatments were assigned randomly. The majority of the patients seem to follow a rather linear evolution however, some also exhibit a nonlinear evolution. It is also apparent that the variability between the patients is larger than the variability within the patients and that there is tracking (patients who start high tend to finish high and likewise for patients who start low). This makes sense since we expect measurements from the same patient to be

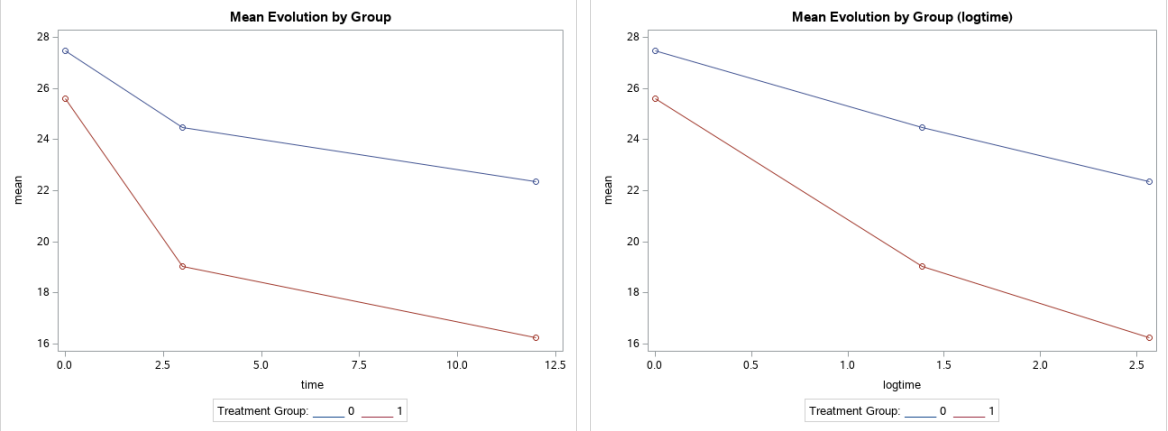


Figure 1: Mean evolution for the outcome variable (headache severity) by group in time scale (left) and logtime scale (right). (group = 0 for placebo, group = 1 for acupuncture).

correlated with each other. The intra-patient correlation needs to be taken into account by our model and this can be accomplished through adding random effects.

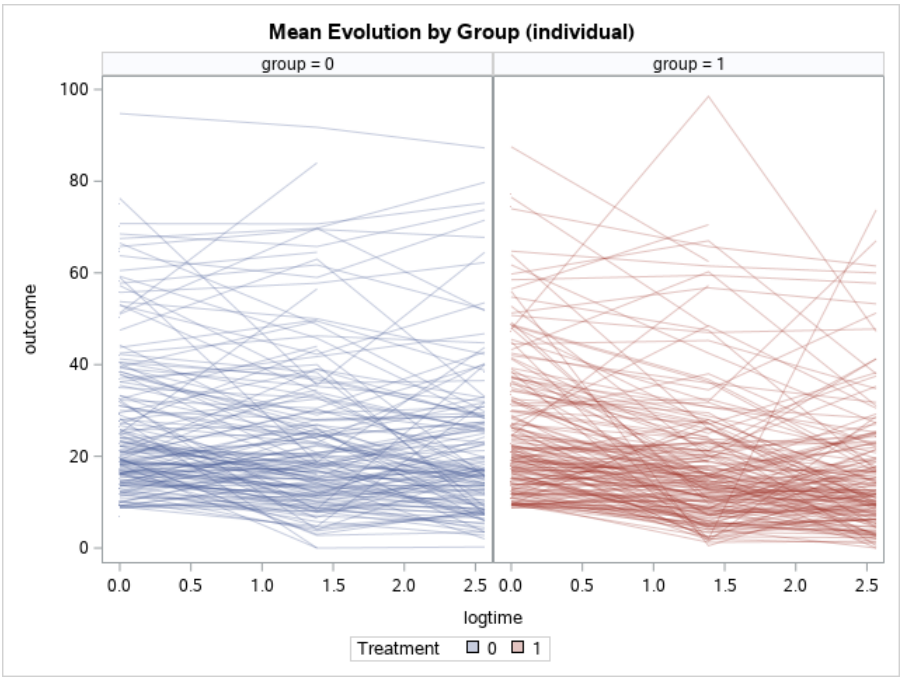


Figure 2: Mean evolution for the outcome variable on logtime scale for each patient, grouped by treatment.

Variance Structure

As seen in Figure 3, for treatment group 0, the variance remains more or less the same when moving from the first measurement to the second and experiences a slight increase in the third. On the other hand, treatment group 1 shows increasing variance in the

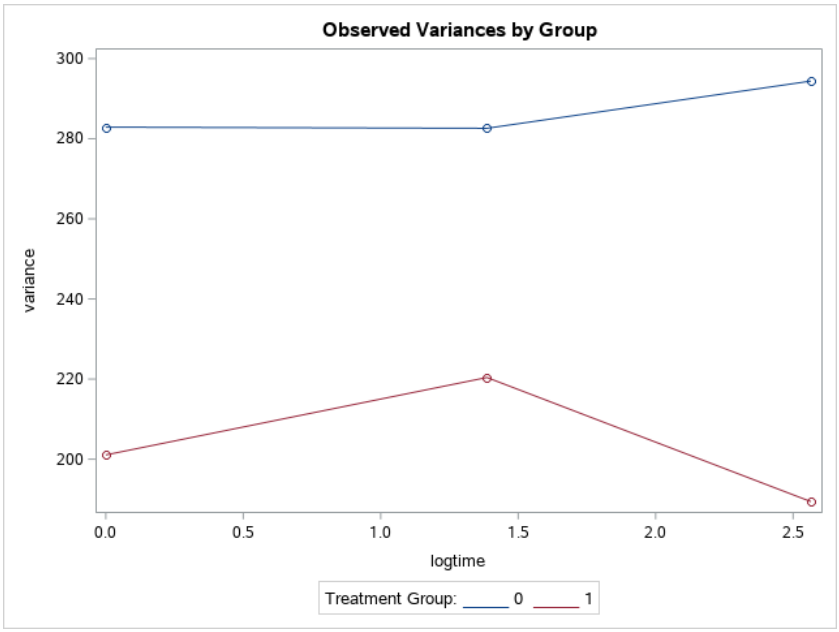


Figure 3: Observed variance over time by treatment group (logtime scale).

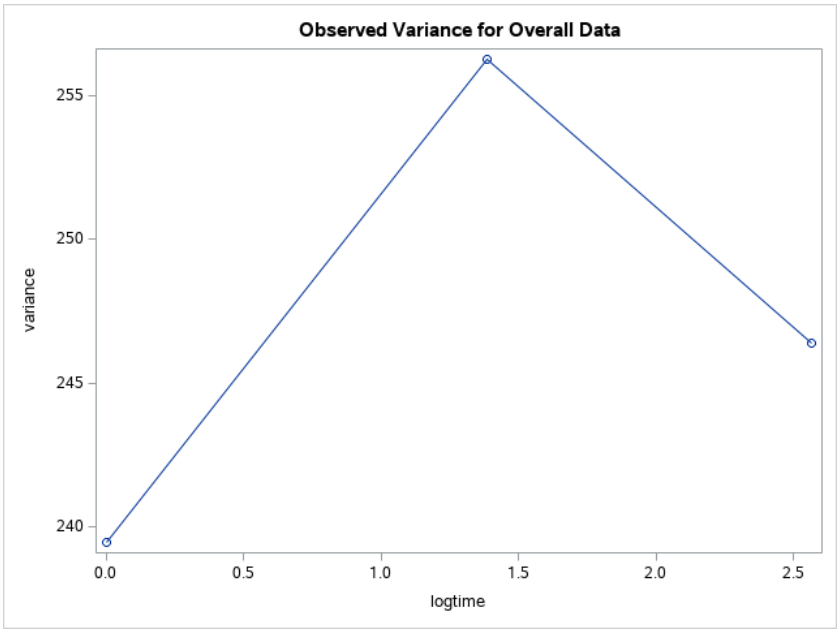


Figure 4: Observed variance over time for entire data set (logtime scale).

second measurement; this variance decreases to lower than baseline level in the third measurement. The two groups follow somewhat opposite variance patterns; our model needs to account for both through one variance function. Certainly, the variance is not constant over time so random effects of *logtime* should be present. Figure 4 of the variance of the overall dataset, showcases a variance which follows a quadratic trend with a negative slope. This can be modelled by adding higher order random effects of *logtime* which allow for a more flexible variance structure.

Correlation Structure

The correlation between measurements taken at baseline is higher with those taken after three months than the ones taken at the end of the study (after 12 months). This pattern is observed in the overall correlation matrix as well as in the group specific matrices. It can be explained by the fact that we expect measurements that are nearer to each other in time to be more highly correlated.

	time_0	**time_3**	**time_12**
time_0	1	0.75	0.72
time_3	0.75	1	0.73
time_12	0.72	0.73	1

Table 2: Correlation structure in the overall dataset.

	group = 0 (placebo)			**group = 1 (acupuncture)**		
	time_0	**time_3**	**time_12**	**time_0**	**time_3**	**time_12**
time_0	1	0.83	0.82	1	0.67	0.58
time_3	0.83	1	0.81	0.67	1	0.62
time_12	0.82	0.81	1	0.58	0.62	1

Table 3: Correlation structure by treatement group.

Linear Mixed Model

We first define the variable $ln(t)$ as:

$$ln(t)_j = ln(time_j + 1)$$

Model 1

To allow for subject-specific intercepts and slopes we fit a linear mixed model which includes a random intercept and a random effect for time. The model is specified as:

$$Y_{ij} = \begin{cases} \beta_1 + b_{1i} + (\beta_2 + b_{2i})ln(t)_j + \epsilon_{ij}, \ \textit{for placebo} \\ \beta_3 + b_{1i} + (\beta_4 + b_{2i})ln(t)_j + \epsilon_{ij}, \ \textit{for acupuncture} \end{cases} \tag{1}$$

with the random effects assumed to follow a joint normal distribution with mean 0 and variance-covariance matrix G, as shown in (2) and (3):

$$(b_{1i}, b_{2i})' \sim N(0, G) \tag{2}$$

$$G = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \tag{3}$$

The model assumes a linear evolution for each subject on the log scale for time with random intercept $\beta_1 + b_{1i}$ or $\beta_3 + b_{1i}$ and random slope $\beta_2 + b_{2i}$ or $\beta_4 + b_{2i}$, corresponding to treatment group = 0 (placebo) and treatment group = 1 (acupuncture).

The implied marginal model is:

$$Y_{ij} = \begin{cases} \beta_1 + \beta_2 ln(t)_j, \ \textit{for placebo treatment} \\ \beta_3 + \beta_4 ln(t)_j, \ \textit{for acupuncture treatment} \end{cases} \tag{4}$$

Equation (4) shows that the average slope is allowed to vary based on the treatment group, which is in line with what we expect from the exploratory analysis (Figure 1).

$$Var(Y_{ij}) = g_{22} ln(t)_j^2 + g_{12} ln(t)_j + g_{11} + \sigma^2_{resid} \tag{5}$$

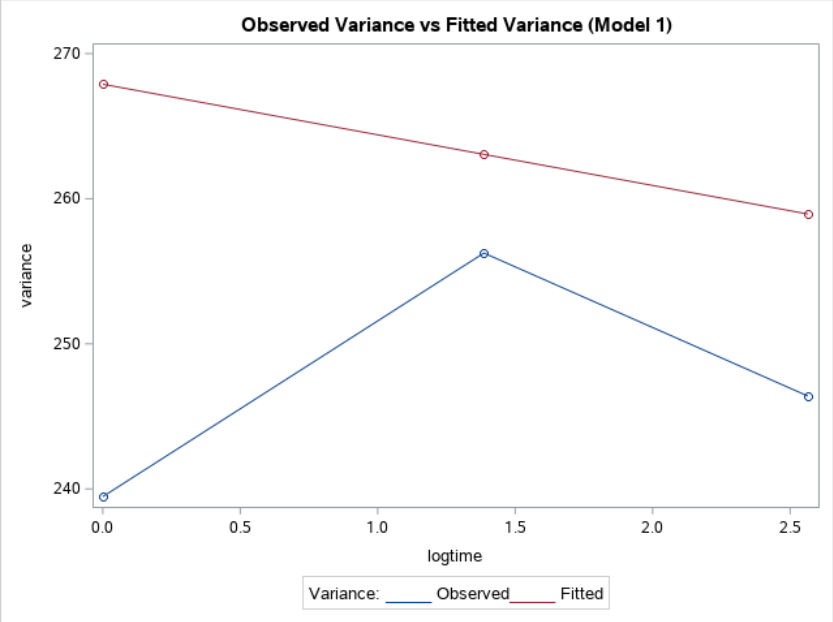


Figure 5: Observed vs fitted variance plot for Model 1.

The implied variance function (Equation (5)) is a quadratic function that depends on $ln(t)_j$. The coefficient in front of the quadratic term is g_{22}, which represents a variance and thus is constrained to be nonnegative. When estimating this model, the resulting G matrix is not positive definite. The reason is that the estimate for g_{22} tries to become negative to capture the shape of the observed variance but because of the constraints it ends up in the boundary of the parameter space (estimated to be 0). Without the off diagonal entries being zero as well, the matrix cannot be positive definite. This result can be seen in Figure 12, where the fitted variance is linear in the log time scale as the quadratic coefficient is estimated to be 0.

Model 2

To remedy the issue with the G matrix, we instead add a squared random effect for $ln(t)$. The specified model becomes:

$$Y_{ij} = \begin{cases} \beta_1 + b_{1i} + \beta_2 ln(t)_j + b_{2i} ln(t)_j^2 + \epsilon_{ij}, \ \textit{for placebo treatment} \\ \beta_3 + b_{1i} + \beta_4 ln(t)_j + b_{2i} ln(t)_j^2 + \epsilon_{ij}, \ \textit{for acupuncture treatment} \end{cases} \tag{6}$$

In the marginal model (refer to Appendix for calculations), the implied variance function changes to a fourth degree polynomial allowing for a more flexible modelling of the variance:

$$Var(Y_{ij}) = g_{22}ln(t)_j^4 + 2ln(t)_j^2 g_{12} + + g_{11} + \sigma^2_{resid} \tag{7}$$

It can be seen in Figure 6 that the shape of the fitted variance function more closely resembles the observed variance:

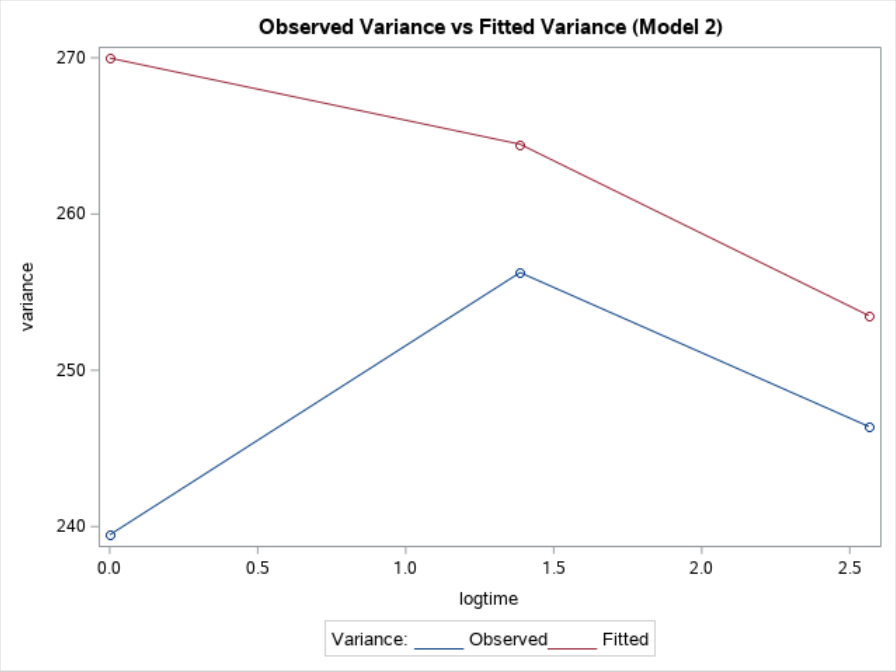


Figure 6: Observed vs fitted variance plot for Model 2.

We fit the model without an intercept. With this parameterization we can directly obtain the estimates for the parameters β_1, β_2, β_3, and β_4, which are shown in Table 4.

Effect	**group**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr** $>\|t\|$
group	0	27.3670	1.1600	293	23.59	<.0001
group	1	25.1673	1.1330	293	22.21	<.0001
logtime*group	0	-1.6974	0.3770	293	-4.50	<.0001
logtime*group	1	-3.3821	0.3551	293	-9.52	<.0001

Table 4: Parameter estimates. The variable logtime is denotes $ln(t)$.

The estimated intercepts for the average of each group are both positive with the one for the acupuncture group being slightly larger, but this difference is not necessarily statistically significant. The slopes for both treatment groups are estimated to be negative, with the one for acupuncture patients being larger in absolute value. This suggests that patients treated with acupuncture have on average a higher reduction rate in headache

severity. However, to test if the difference is significant, we estimate a contrast between the slopes β_2 and β_4, since the hypothesis of interest $H_0 : \beta_2 = \beta_4$ cannot be directly tested from the output. The results of this test are given below in Table 5.

Contrasts				
Label	**Num DF**	**Den DF**	**F Value**	**Pr >F**
placebo-acupuncture	1	293	10.58	0.0013

Table 5: Test for difference in the average evolution between groups.

With a p-value for the F-statistic equal to 0.0013, we conclude that the difference in the slopes is statistically significant ($\alpha = 0.05$), confirming that the reduction rate in headache severity is on average larger for patients treated with acupuncture and pointing towards acupuncture being an effective treatment.

Estimated G Matrix				
Row	**Effect**	**id**	**Col1**	**Col2**
1	Intercept	100	200.94	-1.5101
2	logtime*logtime	100	-1.5101	0.07815

Table 6: Estimated variance-covariance matrix of the random effects.

Table 6 shows the variance-covariance matrix of the random effects estimated by this model. Here we observe that the variance of the random intercept is much larger than that of the squared random effect. This means that most of the variability is accounted for by the difference in intercepts. The covariance between the two random effects is negative. In Table 7 we also see this reflected as a (weak) negative correlation. The interpretation is that, patients with an initial higher level of headache severity have a more negative slope (more reduction in headache severity).

Estimated G Correlation Matrix				
Row	**Effect**	**id**	**Col1**	**Col2**
1	Intercept	100	1.0000	-0.3801
2	logtime*logtime	100	-0.3801	1.0000

Table 7: Estimated correlation matrix of the random effects.

Lastly, we consider the fitted variance-covariance and correlation matrices for a particular subject. In Table 8, we see that the variance decreases over time. This pattern does not quite match the observed pattern in the overall dataset. In Table 9, the fitted correlation matrix is shown. Here, the correlation pattern is close to the observed one where observations are slightly more correlated with observations that are nearer in time.

	time_0	**time_3**	**time_12**
time_0	269.98	198.04	191.00
time_3	198.04	264.47	189.09
time_12	191.00	189.09	253.50

Table 8: Estimated variance-covariance matrix for subject with id = 141.

	time_0	**time_3**	**time_12**
time_0	1	0.7411	0.7301
time_3	0.7411	1	0.7303
time_12	0.7301	0.7303	1

Table 9: Estimated correlation matrix for subject with id = 141.

Random Effects

The random intercepts and random slopes, estimated by Empirical Bayes estimates, permit the estimation of subject specific intercepts and slopes. A scatter plot of the slopes versus the intercepts (Figure 7), reveals that subjects with higher intercepts and therefore higher initial headache scores have more negative slopes, as evidenced by the negative linear relationship between the slopes and the intercepts. This negative linear trend appears in both the placebo group as well as the acupuncture group, and is consistent with the estimated covariance between the random intercept and the $logtime * logtime$ random effect. Moreover, if the initial headache scores are high, then the reduction rate of the headache scores increases over time.

Additionally, the histograms for the random intercepts and slopes (Figure 8) corroborate the larger variance for the intercepts than for the random slopes initially indicated by the estimated G matrix, with the means for both distributions around 0, also consistent with the assumption that the expected value of the random effects is in fact 0.

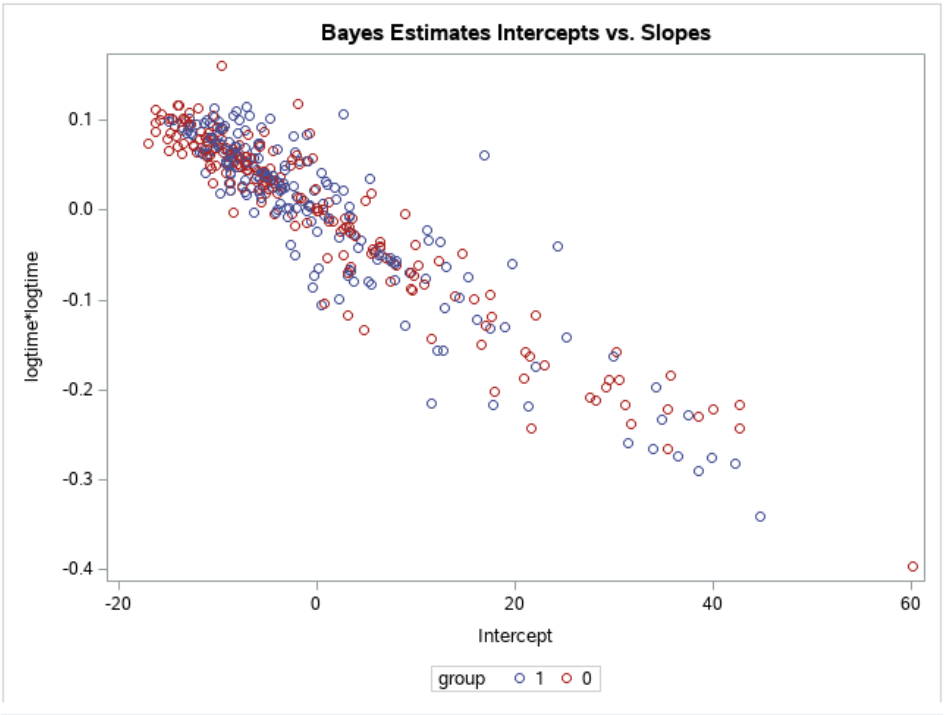
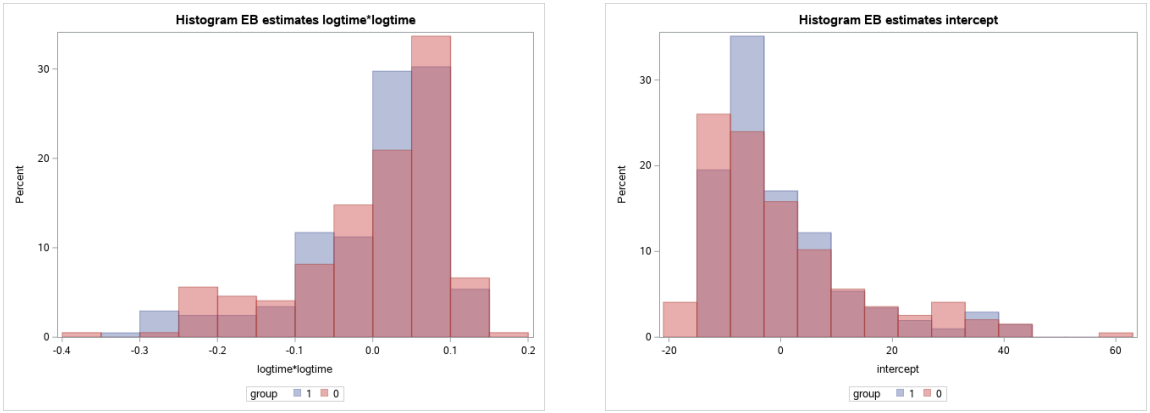


Figure 7: Scatterplot of random intercepts vs. random slopes.



(a) Random Slopes

(b) Random Intercepts

Figure 8: Empirical Bayes Estimates histograms.

Adding Covariates to the Model

In order to correct for baseline differences between subjects, we include the covariates *frequency*, *age*, and *sex*. Despite many attempts at respecifying the model, we were unable to find any remedy for the issue of the estimated G matrix being not positive definite. While acknowledging the issues surrounding such a problem, we choose not to remove the random effects initially specified in Model 2 so the two models remain comparable. The new model is specified as:

$$Y_{ij} = \begin{cases} \beta_1 + b_{1i} + \beta_2 ln(t)_j + \beta_3 age_j + \beta_4 freq_j + \beta_5 sex_j + b_{2i} ln(t)_j^2 + \epsilon_{ij} \text{ for placebo} \\ \beta_6 + b_{1i} + \beta_7 ln(t)_j + \beta_3 age_j + \beta_4 freq_j + \beta_5 sex_j + b_{2i} ln(t)_j^2 + \epsilon_{ij} \text{ for acupuncture} \end{cases} \tag{8}$$

In the final model with baseline covariates, we observe statistically significant effects on headache scores for *gender*, *age*, and $frequency$ (Table 10). In terms of the fixed effects, the intercepts for the placebo and acupuncture groups become negative, with values of -11.91 and -13.40 respectively. Being female leads to a 3.93 increase in headache scores on average, and each additional year older an individual is leads to a .131 increase in headache scores; for each additional day with a headache, participants record a 1.82 increase in headache scores.

Effect	**Estimates**	**Std.Error**	**Pr>|t|**
placebo	-11.91(27.37)	2.83 (1.16)	<.0001 (<.0001)
acupuncture	-13.40 (25.17)	2.82(1.13)	<.0001 (<.0001)
logtime*group 0	-1.66 (-1.70)	0.38 (0.38)	<.0001 (<.0001)
logtime*group1	-3.34 (-3.38)	0.36 (0.35)	<.0001 (<.0001)
sex	3.93	1.32	.0031
age	0.13	0.04	.0029
frequency	1.82	0.07	<.0001

Table 10: Parameter estimates for model with additional covariates. Estimates for model without baseline covariates are shown in brackets.

Both the placebo and the acupuncture group still show a significant headache score reduction over time, with the acupuncture group showing a larger reduction in headache scores over time when compared to the placebo group, as shown in Table 11.

Difference Slopes				
Label	**Estimate**	**Den DF**	**t Value**	**Pr >|t|**
placebo-acupuncture	1.6750	293	3.22	0.0014

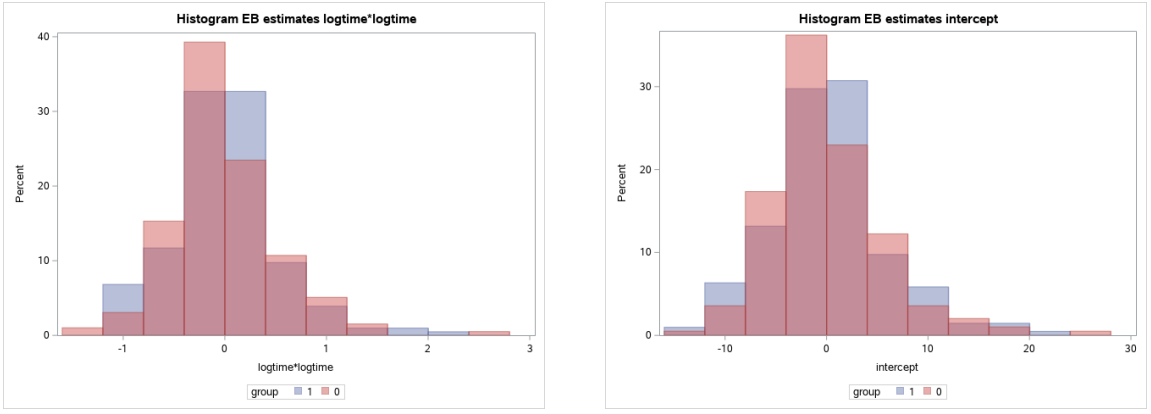
Table 11: Test for difference in the average evolution between groups.

As seen in the estimated G matrix for the expanded model (Table 12), we observe a

significant decrease in the variance of the random intercepts, now 44.53 with the model with baseline covariates. Now, the variances of the random intercepts and random slopes account for 39.69% of the total variance. The decrease in the variance of the random intercepts, which is seen in the below histograms of the Bayes estimates for the random slopes (Figure 9), are likely a reflection of the additional baseline covariates *sex*, *age*, and *frequency* partially explaining some of the variance of the random intercepts. Additionally, the covariance between the random slopes and intercepts now has a positive association indicated by the positive covariance of 4.88. This implies that participants with higher initial headache scores have lower headache reduction rates. This is further corroborated by the positive linear trend evidenced by the scatterplot between the Bayes estimates for the random intercepts and random slopes (Figure 10).

Estimated G Matrix Expanded Model				
Row	**Effect**	**id**	**Col1**	**Col2**
1	Intercept	100	44.5259	4.8753
2	logtime*logtime	100	4.8753	0.1032

Table 12: Estimated G matrix of the random effects for expanded model.



(a) Random Slopes

(b) Random Intercepts

Figure 9: Empirical Bayes Estimates histograms for the expanded model.

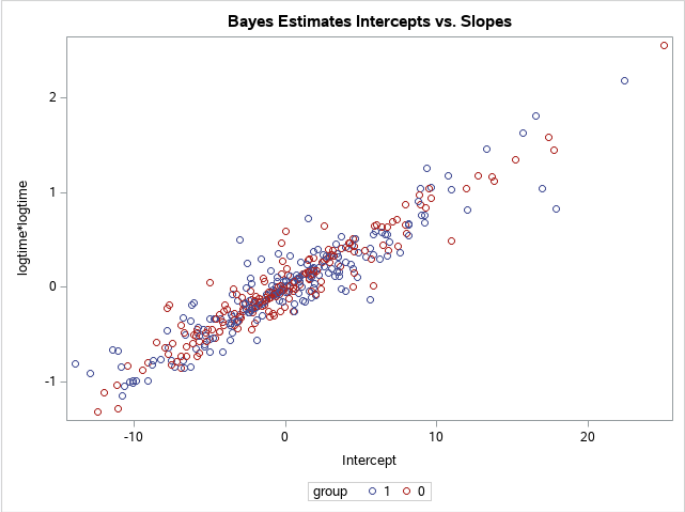


Figure 10: Scatterplot of random intercepts vs. random slopes for expanded model.

Logistic Mixed Model

Dichotomization of the Outcome Variable

Continuous variables can be dichotomized as a way to categorize patients into groups. In this section, we will transform the outcome scores into two categories and the data will be re-analyzed via a logistic mixed model.

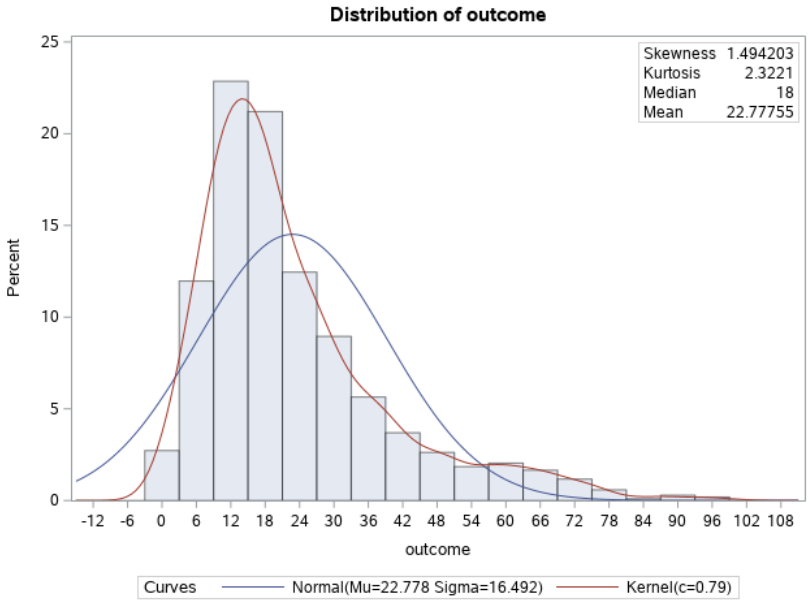


Figure 11: Distribution of outcome

The median of the overall patient sample is 18. We can say that 18 might be a good cutoff

value and the median has been reported as a statistically sound technique for choosing an optimal cutoff point for dichotomization. The observations whose outcome score is equal or larger to 18 are categorized as high severity group with value 1, and those whose score is lower than 18 are categorized into low severity group with value 0.

Fitting the Logistic Mixed Model

It is then assumed that the new outcome variable follows Bernoulli distribution with parameter π_{ij}, which gives the probability of an observation belonging to the high severity group:

$$Newoutcome \sim Bernoulli(\pi_{ij}) \tag{9}$$

We fit the model with random intercepts and random slopes. In this model, the effects of age, sex and frequency are also explored. The estimated G matrix is:

$$\mathbf{G} = \begin{pmatrix} 0 & 0.1765 \\ 0.1765 & 0 \end{pmatrix} \tag{10}$$

We note that for this logistic mixed model with random intercept and slopes, the G matrix is not positive definite, but we decided to keep it in order to compare with the previous models. We will still be able to make valid inferences on the fixed effects.
The model can be expressed as follows: π_{ij} refers to the probability of selecting observations into the high severity group and the subscript ij denotes the jth measurement on ith patient. The link function is the *logit* function and the mean can be modelled by the logarithm of odd ratios between the probability of selecting patients into the high severity group and the probability of selecting patients into the low severity group. β_1 and β_6 refer to the group specific intercepts and they are assumed to follow normal distribution with two parameters, mean (0) and variance (σ_1^2, σ_6^2). β_2 and β_7 refer to the group specific slopes and are also assumed to follow a normal distribution with two parameters, mean (0) and variance (σ_1^2, σ_6^7). β_4, β_5 and β_6 represent the effects of age, $frequency$ and sex respectively.

$$logit(\pi_{ij}) = \begin{cases} \beta_1 + b_{1i} + \beta_2 ln(t)_j + \beta_3 age_j + \beta_4 freq_j + \beta_5 sex_j + b_{2i} ln(t)_j^2 + \epsilon_{ij,} \text{ for placebo} \\ \beta_6 + b_{1i} + \beta_7 ln(t)_j + \beta_3 age_j + \beta_4 freq_j + \beta_5 sex_j + b_{2i} ln(t)_j^2 + \epsilon_{ij}, \text{ for acupuncture} \end{cases} \tag{11}$$

The model above is estimated by the method PQL (Penalized Quasi-Likelihood) and the estimation converges. The table below gives the estimation result of the model:

Effect	**Estimates**	**Std Error**	**Pr>|t|**
group placebo	-3.5720	0.5226	<.0001
group treatment	-3.7828	0.5283	<.0001
age	0.01046	0.007675	0.1738
sex	0.7927	0.2396	0.001
frequency	0.1982	0.01467	<.0001
logtime*group (placebo)	-0.7609	0.1024	<.0001
logtime*group (treatment)	-0.9095	0.1042	<.0001

Table 13: Parameter estimates for logistic mixed model.

The effect of age is non-significant in our model, implying that *age* does not have an impact on the severity of headaches of patients. On the other hand, *sex* and $frequency$ are both statistically significant. The coefficients of the logistic model can be interpreted as follows: the exponential of β_5 is equal to 2.209. It indicates that when other variables are controlled, the odds of a patient in the high severity group being female rather than male are 2.209 times higher. A similar interpretation can be given for the other parameters.

Next we look at the effects of the treatment groups and time.

Estimates					
Label	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr >|t|**
difference slopes	-0.1487	0.1436	399	-1.04	0.3012

Table 14: Comparison of slopes between treatment groups for logistic mixed model.

The estimation of β_1 - β_6 indicates the difference between the intercepts of the two treatment groups. The results show that the probability of acupuncture treatment group patients falling into the high severity group is 0.81 times $(exp(-3.7828 + 3.5720))$ the probability of placebo group patients being categorized as high severity group. Also, the effect of $logtime * group$ indicates that as time progresses, people from the treatment group are less likely to fall under the high severity group. Indeed, the treatment group specific slope is lower than the placebo group specific slope (-0.9095 against -0.7608).

However, there is no statistically significant difference between these 2 slopes as shown in Table 14. This result differs from the one observed in the expanded linear mixed model. This is confirmed by the plot of marginal average evolution, with A corresponding to placebo group and B corresponding to the acupuncture treatment group.

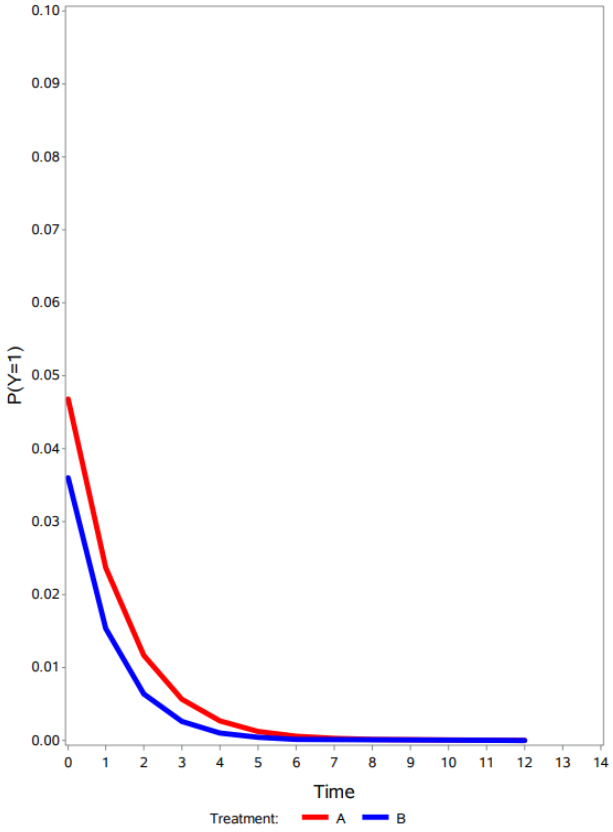


Figure 12: Marginal Average Evolution of π_{ij}

The use of a logistic mixed model here allowed for a binary classification of patients into groups defined by 'high severity' and 'low severity'. With this dichotomization, we created a model structure that could be used to aid healthcare practitioners to refer patients to medical services most appropriate for them based on their age, sex and frequency of headaches. For instance, female patients have higher odds of being classified as high severity of headaches. However, the insignificance of factor estimates observed for $logtime * group$ and age in the logistic model should be interpreted with caution because of loss of information typically associated with logistic models. Such loss of information can render previously significant factor effects insignificant, as shown in the linear mixed model in the previous section.

Conclusion

Upon fitting a linear mixed model with subject specific intercepts and a quadratic logtime random effect, we found that subjects with a high initial headache score experience a higher headache score reduction rate than those with lower initial headache scores. Moreover, subjects in the acupuncture group experienced a more robust reduction rate in headache scores over time compared to the placebo group. When incorporating the baseline characteristics via covariates age, $frequency$, and sex, we observed that women had significantly higher headache scores then men, and that age and each additional day with a headache led to significantly higher headache scores as well. Furthermore, headache scores decreased over time for both groups, although the acupuncture group experienced a faster reduction rate than the placebo group with respect to headache scores. Most importantly, the inclusion of baseline covariates resulted in a reversal of the covariance between subject specific intercepts in slopes, whereby subjects with a higher initial headache score experience a lower headache score reduction rate than those with lower initial headache scores.

When fitting a logistic linear mixed model with subject specific intercepts and a quadratic logtime random effect, it was established that patients of the high severity group have higher odds of being female patients than male patients and higher odds of having more frequent headaches than less frequent headaches. However, results showed that we don't have enough evidence to support the effect of group on the evolution of the severity of headaches. It can be explained by the fact that the logistic mixed model results in loss of information caused by the dichotomization of outcomes into two groups. Based on the factor effect estimates of the linear mixed model, we believe that the type of treatment does have a significant impact on the evolution of the outcomes of patients.

References

Vickers, Andrew J., et al. "Acupuncture for chronic headache in primary care: large, pragmatic, randomised trial." *Bmj* 328.7442 (2004): 744.

Appendix

Calculations for Marginal Model

For Model 2, specified as:

$$Y_{ij} = \begin{cases} \beta_1 + b_{1i} + \beta_2 ln(t)_j + b_{2i} ln(t)_j^2 + \epsilon_{ij}, \ \textit{for placebo treatment} \\ \beta_3 + b_{1i} + \beta_4 ln(t)_j + b_{2i} ln(t)_j^2 + \epsilon_{ij}, \ \textit{for acupuncture treatment} \end{cases} \tag{12}$$

The expected value:

Proof.

$$E[Y_{ij}] = \begin{cases} E[\beta_1 + b_{1i} + \beta_2 ln(t)_j + b_{2i} ln(t)_j^2 + \epsilon_{ij}] \\ E[\beta_3 + b_{1i} + \beta_4 ln(t)_j + b_{2i} ln(t)_j^2 + \epsilon_{ij}] \end{cases} \tag{13}$$

$$\implies E[Y_{ij}] = \begin{cases} E[\beta_1] + E[b_{1i}] + E[\beta_2 ln(t)_j] + E[b_{2i} ln(t)_j^2] + E[\epsilon_{ij}] \\ E[\beta_3] + E[b_{1i}] + E[\beta_4 ln(t)_j] + E[b_{2i} ln(t)_j^2] + E[\epsilon_{ij}] \end{cases} \tag{14}$$

$$\implies E[Y_{ij}] = \begin{cases} \beta_1 + 0 + \beta_2 ln(t)_j + ln(t)_j^2 E[b_{2i}] + 0 \\ \beta_3 + 0 + \beta_4 ln(t)_j + ln(t)_j^2 E[b_{2i}] + 0 \end{cases} \tag{15}$$

$$\implies E[Y_{ij}] = \begin{cases} \beta_1 + 0 + \beta_2 ln(t)_j + ln(t)_j^2 * 0 + 0 \\ \beta_3 + 0 + \beta_4 ln(t)_j + ln(t)_j^2 * 0 + 0 \end{cases} \tag{16}$$

$$\implies E[Y_{ij}] = \begin{cases} \beta_1 + \beta_2 ln(t)_j \\ \beta_3 + \beta_4 ln(t)_j \end{cases} \tag{17}$$

□

The variance:

Proof.

$$Var(Y_{ij}) = Var(\beta_1 P + \beta_3 A + b_{1i} + \beta_2 ln(t)_j P + \beta_4 ln(t)_j A + b_{2i} ln(t)_j^2 + \epsilon_{ij}) \tag{18}$$

where,

$$P = \begin{cases} 0, \ \textit{if acupuncture treatment} \\ 1, \ \textit{if placebo treatment} \end{cases}$$

$$A = \begin{cases} 0, & \textit{if placebo treatment} \\ 1, & \textit{if acupuncture treatment} \end{cases}$$

$$\begin{aligned} Var[Y_{ij}] = Var[\beta_1 P] + Var[\beta_3 A] + Var[b_{1i}] + Var[\beta_2 ln(t)_j P] + Var[\beta_4 ln(t)_j A] \\ + Var[b_{2i} ln(t)_j^2] + Var[\epsilon_{ij}] + 2Cov[b_{1i}, b_{2i} ln(t)_j^2] \end{aligned} \tag{19}$$

$$Var(Y_{ij}) = 0 + 0 + g_{11} + 0 + 0 + ln(t)_j^4 g_{22} + \sigma^2_{resid} + 2ln(t)_j^2 g_{12} \tag{20}$$

$$Var(Y_{ij}) = ln(t)_j^4 g_{22} + 2ln(t)_j^2 g_{12} + g_{11} + \sigma^2_{resid} \tag{21}$$

□

Similar calculations can be carried out for Model 1 as well.

SAS Code

```
data acupuncture;
set "/home/u49834661/KUL/acupuncture.sas7bdat";
logtime = log(time + 1);
run;

/* QUESTION 1 */
title "Exploratory analysis";
proc summary data = acupuncture nway;
class group time;
var outcome;
output mean = mean nmiss = nmiss;
proc print;
run;

/* Mean Structure */
proc sort data = acupuncture;
by group;
run;
```

```
proc means data = acupuncture nway;
class time group logtime;
var outcome;
output out = meansout mean = mean;
run;

proc sgplot data=meansout;
   scatter x=time y=mean /group = group;
   series x=time y=mean / group=group grouplc=group lineattrs=(color=blue);
   keylegend 'grouping' / type = linecolor;
   title 'Mean Evolution by Group';
   /* Use a FOOTNOTE statement to simulate a legend. */
   footnote1 box=1 bcolor=white
             'Treatment Group: '
             color=VIGB "_____"
             color=black " 0 "
             color=depk  "_____"
             color=black " 1";
run;

proc sgplot data=meansout;
   scatter x=logtime y=mean /group = group;
   series x=logtime y=mean / group=group grouplc=group lineattrs=(color=blue);
   keylegend 'grouping' / type = linecolor;
   title 'Mean Evolution by Group (logtime)';
   /* Use a FOOTNOTE statement to simulate a legend. */
   footnote1 box=1 bcolor=white
             'Treatment Group: '
             color=VIGB "_____"
             color=black " 0 "
             color=depk  "_____"
             color=black " 1";
run;
footnote1;
proc sgpanel data = acupuncture;
title 'Mean Evolution by Group (individual)';
panelby group / columns = 2 onepanel sparse;
series x = logtime y = outcome / group = id break transparency= 0.7 grouplc =
    group lineattrs=(pattern=solid);
keylegend / type = linecolor title = "Treatment";
run;

/* Variance and correlation structure */
```

```
data time_0_0 (rename=(outcome = outcome_0))
    time_0_3 (rename=(outcome = outcome_3))
    time_0_12 (rename=(outcome = outcome_12))
    time_1_0 (rename=(outcome = outcome_0))
    time_1_3 (rename=(outcome = outcome_3))
    time_1_12 (rename=(outcome = outcome_12));
set acupuncture;
if group = 0 AND time = 0 then output time_0_0;
else if group = 0 AND time = 3 then output time_0_3;
else if group = 0 AND time = 12 then output time_0_12;
else if group = 1 AND time = 0 then output time_1_0;
else if group = 1 AND time = 3 then output time_1_3;
else if group = 1 AND time = 12 then output time_1_12;
run;

data time_0 (rename=(outcome = outcome_0))
    time_3 (rename=(outcome = outcome_3))
    time_12 (rename=(outcome = outcome_12));
set acupuncture;
if time = 0 then output time_0;
else if time = 3 then output time_3;
else if time = 12 then output time_12;
run;

proc sort data = time_0;
by id;
run;
proc sort data = time_3;
by id;
run;
proc sort data = time_12;
by id;
run;

data acupuncture_g0;
merge time_0_0 time_0_3 time_0_12;
by id;
run;

data acupuncture_g1;
merge time_1_0 time_1_3 time_1_12;
by id;
run;
```

```
data acupuncture_var;
merge time_0 time_3 time_12;
by id;
run;

title "Variance-Covariance Matrix for Placebo Group (group = 0)";
proc corr data = acupuncture_g0 noprob out = g0_cov nomiss nosimple nocorr
    cov ;
var outcome_0 outcome_3 outcome_12;
run;

title "Variance-Covariance Matrix for Acupuncture Group (group = 1)";
proc corr data = acupuncture_g1 noprob out = g1_cov nomiss nosimple nocorr
    cov;
var outcome_0 outcome_3 outcome_12;
run;

title "Variance-Covariance Matrix for Overall Data";
proc corr data = acupuncture_var noprob out = cov nomiss nosimple nocorr cov;
var outcome_0 outcome_3 outcome_12;
run;

title "Correlation Matrix for Placebo Group (group = 0)";
proc corr data = acupuncture_g0 noprob out = g0_corr nomiss nosimple;
var outcome_0 outcome_3 outcome_12;
run;

title "Correlation Matrix for Acupuncture Group (group = 1)";
proc corr data = acupuncture_g1 noprob out = g1_corr nomiss nosimple;
var outcome_0 outcome_3 outcome_12;
run;

title "Correlation Matrix for Overall Data";
proc corr data = acupuncture_var noprob out = corr nomiss nosimple;
var outcome_0 outcome_3 outcome_12;
run;

proc iml;
use acupuncture(obs = 3);
read all var {time logtime} into M;
create times var{"time" "logtime"};
append from M;
```

```
close;
title;

proc iml;
use g0_cov(obs = 3);
read all var {outcome_0 outcome_3 outcome_12} into M;
X = vecdiag(M);
create g0_var var{"g0"};
append from X;
close;

proc iml;
use g1_cov(obs = 3);
read all var {outcome_0 outcome_3 outcome_12} into M;
X = vecdiag(M);
create g1_var var{"g1"};
append from X;
close;

proc iml;
use cov(obs = 3);
read all var {outcome_0 outcome_3 outcome_12} into M;
X = vecdiag(M);
create ov_variance var{"both"};
append from X;
close;

data variances;
merge g0_var g1_var times;
input time;
datalines;
0
3
12
;
run;

data overall_var;
set ov_variance;
input time;
datalines;
0
3
```

```
12
;
run;

data overall_var;
merge overall_var times;
by time;
run;

proc sgplot data=variances noautolegend;
   scatter x = logtime y = g0 / markerattrs = (color = VIGB);
   scatter x = logtime y = g1 /markerattrs = (color = depk);
   series x = logtime y = g0 /lineattrs = (color = VIGB);
   series x = logtime y = g1 /lineattrs = (color = depk);
   yaxis label = "variance";
   title "Observed Variances by Group";
   footnote1 box=1 bcolor=white
             'Treatment Group: '
             color=VIGB "_____"
             color=black " 0 "
             color=depk  "_____"
             color=black " 1";
run;
footnote1;

proc sgplot data = overall_var noautolegend;
   scatter x = logtime y = both;
   series x = logtime y = both;
   yaxis label = "variance";
   title "Observed Variance for Overall Data";
run;


/* QUESTION 2 */
ods output V = m1_V;
proc mixed data = acupuncture;
class id group(ref = "0");
model outcome = group logtime*group/ noint solution;
random intercept logtime / type = un subject = id v = 16 vcorr = 16 g gcorr;
run;

proc iml;
use m1_V(obs = 3);
```

```
read all var {Col1 Col2 Col3} into M;
X = vecdiag(M);
create m1_variance var{"modelvar"};
append from X;
close;

data m1_variance;
set m1_variance;
input time;
datalines;
0
3
12
;
run;

data m1_var_plot;
merge m1_variance overall_var times;
by time;
run;

proc sgplot data = m1_var_plot noautolegend;
   scatter x = logtime y = both / markerattrs = (color = VIGB);
   scatter x = logtime y = modelvar /markerattrs = (color = depk);
   series x = logtime y = both /lineattrs = (color = VIGB);
   series x = logtime y = modelvar /lineattrs = (color = depk);
   yaxis label = "variance";
   title "Observed Variance vs Fitted Variance (Model 1)";
   footnote1 box=1 bcolor=white
             'Variance: '
             color=VIGB "_____"
             color=black " Observed"
             color=depk  "_____"
             color=black " Fitted";
run;
footnote1;

/* MODEL 2 */
ods output V = m2_V;
proc mixed data = acupuncture;
class id group;
model outcome = group logtime*group/ noint solution;
random intercept logtime*logtime / type = un subject = id v = 16 vcorr = 16 g
```

```
    gcorr;
contrast 'placebo-acupuncture(sl)' logtime*group 1 -1;
run;

proc iml;
use m2_V(obs = 3);
read all var {Col1 Col2 Col3} into M;
X = vecdiag(M);
create m2_variance var{"modelvar"};
append from X;
close;

data m2_variance;
set m2_variance;
input time;
datalines;
0
3
12
;
run;

data m2_var_plot;
merge m2_variance overall_var times;
by time;
run;

proc sgplot data = m2_var_plot noautolegend;
   scatter x = logtime y = both / markerattrs = (color = VIGB);
   scatter x = logtime y = modelvar /markerattrs = (color = depk);
   series x = logtime y = both /lineattrs = (color = VIGB);
   series x = logtime y = modelvar /lineattrs = (color = depk);
   yaxis label = "variance";
   title "Observed Variance vs Fitted Variance (Model 2)";
   footnote1 box=1 bcolor=white
             'Variance: '
             color=VIGB "_____"
             color=black " Observed"
             color=depk  "_____"
             color=black " Fitted";
run;
footnote1;
```

```
*Question 3:

Model with Random Effects Bayes Estimates;
proc mixed data= acupuncture;
class id group;
   model outcome = group group*logtime  /noint solution;
   random intercept logtime*logtime / type=un subject= id g gcorr v vcorr=16
       solution;
   estimate 'difference slopes' group*logtime 1 -1;
   ods output solutionr=bayes;
   run;

*Plotting the Bayes estimates for Intercepts vs. Slopes;
data acupuncture_groups;
set acupuncture;
keep id group;
run;

proc sort data = acupuncture_groups dupout = groups_nodup nodupkey;
by id;
run;

data bayes;
merge bayes groups_nodup;
by id;
run;

data bayesplot;
   set bayes;
   by id;
   retain Intercept Logtime2 id group;
   keep Intercept Logtime2 id group;
   if Effect="Intercept" then Intercept=Estimate;
   else if Effect="logtime*logtime" then Logtime2=Estimate;
   if last.id=1 then output;
   run;

proc sgplot data=bayesplot;
 scatter x=Intercept y=Logtime2 / group=group;
  title "Bayes Estimates Intercepts vs. Slopes";
  yaxis label = "logtime*logtime";
```

```
run;

*histogram for EB estimates question 3;
proc sgplot data=bayesplot;
   histogram logtime2 / group= group transparency=0.5 ;
   xaxis label= "logtime*logtime";
   title "Histogram EB estimates logtime*logtime";
run;
proc sgplot data=bayesplot;
   histogram intercept / group= group transparency=0.5 ;
   xaxis label= "intercept";
   title "Histogram EB estimates intercept";
run;


*Question 4:;
proc mixed data= acupuncture;
class id group;
   model outcome = sex age frequency group group*logtime / noint solution;
   random intercept logtime*logtime / type=un subject= id g gcorr v=16
       vcorr=16 solution;
   estimate 'difference slopes' group*logtime 1 -1;
   ods output solutionr=bayes2;
   run;


*Plotting the Bayes estimates for Intercepts vs. Slopes;

data acupuncture_groups;
set acupuncture;
keep id group;
run;

proc sort data = acupuncture_groups dupout = groups_nodup nodupkey;
by id;
run;

data bayes2;
merge bayes2 groups_nodup;
by id;
run;

data bayesplot2;
```

```
   set bayes2;
   by id;
   retain Intercept Logtime2 id group;
   keep Intercept Logtime2 id group;
   if Effect="Intercept" then Intercept=Estimate;
   else if Effect="logtime*logtime" then Logtime2=Estimate;
   if last.id=1 then output;
   run;

proc sgplot data=bayesplot2;
 scatter x=intercept y=Logtime2 / group=group;
 title "Bayes Estimates Intercepts vs. Slopes";
  yaxis label = "logtime*logtime";
run;

*histogram for EB estimates question 4;
proc sgplot data=bayesplot2;
   histogram logtime2 / group= group transparency=0.5;
   xaxis label= "logtime*logtime";
   title "Histogram EB estimates logtime*logtime";
run;
proc sgplot data=bayesplot2;
   histogram intercept / group= group transparency=0.5 ;
   xaxis label= "intercept";
   title "Histogram EB estimates intercept";
run;



/* Question 5 */
/*Find Outcome median*/
proc univariate data= acupuncture;
var outcome;
histogram outcome / normal (mu=est sigma= est) kernel;
inset skewness kurtosis median mean / position= ne;
probplot outcome / normal (mu=est sigma= est);
inset skewness kurtosis;
title 'Outcome Distribution';
run;

/*Dichotomization of outcome*/
Data acupuncture2;
Set acupuncture;
```

```
If outcome => 18 then newout=1;
If outcome <18 then newout=0;
run;
proc freq data=acupuncture2;
table newout;
run;

/* Question 6 */
/* random intercept and slope logistic model */
proc glimmix data=acupuncture2 method=rspl;
class group(ref="0") id;
model newout (event='1') = sex frequency age group group*logtime / noint
    dist=binary solution ;
random intercept logtime*logtime / type=un subject=id g gcorr v=16 vcorr=16;
estimate 'difference slopes' group*logtime 1 -1;
run;

/*marginal evolution*/

data h;
   do group=0 to 1 by 1;
      do id=100 to 912 by 1;
         b=sqrt(1.3224)*rannor(-1) ;
            do time=0 to 12 by 1;
   if group=0 then newoutcome=exp(-3.5720 + b -0.7609*time)/(1+ exp(-3.5720 +
       b -0.7609*time));
else newout=exp(-3.7828 + b -0.905*time)/(1+ exp(-3.7828 + b -0.905*time));
output;
end;
end;
end;

proc sort data=h;
by time group;
run;
proc means data=h;
var newout;
by time group;
output out=out;
run;

proc gplot data=out;
plot newout*time=group / haxis=axis1 vaxis=axis2 legend=legend1;
```

```
axis1 label=(h=2 'Time') value=(h=1.5) order=(0 to 14 by 1) minor=none;
axis2 label=(h=2 A=90 'P(Y=1)') value=(h=1.5) order=(0 to 0.1 by 0.01)
   minor=none;
legend1 label=(h=1.5 'Treatment: ') value=(h=1.5 'A' 'B');
title h=2.5 ' Marginal average evolutions (GLMM)';
symbol1 c=red i=join w=10 l=1 mode=include;
symbol2 c=blue i=join w=10 l=1 mode=include;
where _stat_='MEAN';
run;
```