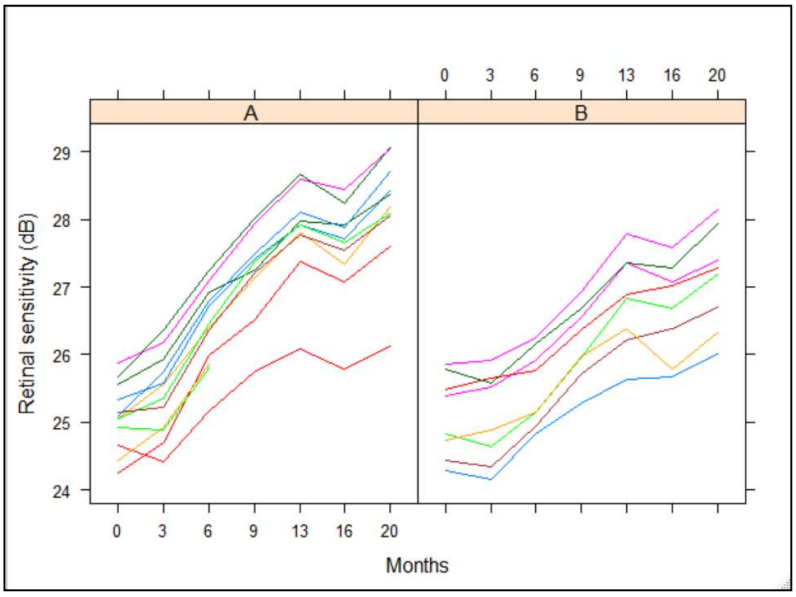
## Exercise 2 Long-term follow-up of chronic central serous chorioretinopathy after successful treatment with photodynamic therapy or micropulse laser

**Answer parts a, b, c, g, h, j, and k, but choose one from parts d1, d2, one from e1, e2, one from f1, f2, and one from i1, i2. So, e.g. d1, e1, f2, i2 is fine, but e.g. d1, e1, f1, f2, i1 is not, because then f2 will not be graded.**

A prospective long-term follow-up study has been set up to study the long-term effects of half-dose photodynamic therapy and high-density subthreshold micropulse laser in patients with chronic central serous chorioretinopathy. The investigators followed up 20 patients: 12 in group A and 8 in group B. The level of the retinal sensitivity on microperimetry (continuous variable) and presence of subretinal fluid (binary variable) was measured 8 months after initiation of treatment (day 0) and at 3, 6, 9, 13, 16 and 20 months after initiation of treatment.

Due to side-effects on month 6, 2 patients underwent the therapy again and thus no data are available for them after month 6.

The data collected for retinal sensitivity are shown in the following plots:




The proportion of patients with presence of subretinal fluid over time and per group is:

| Month | A | B |
|---|---|---|
| 0 | 0.67 | 0.5 |
| 3 | 0.67 | 0.5 |
| 6 | 1 | 0.75 |
| 9 | 1 | 1 |
| 13 | 1 | 1 |
| 16 | 1 | 1 |
| 20 | 1 | 1 |

The investigators started the analyses with the continuous outcome, namely the retinal sensitivity. They fitted two models: Model A and Model B. For the mean part they assumed the same saturated model for both Model A and Model B. For the random part, they considered a different option per model. The results from both models are given in the output in Appendix I below at the end of this exercise. The variable names which appear in the output are:

- ID: patient number,
- Treat: group A or B,
- Y: retinal sensitivity (numeric),
- Fluid: presence of fluid: 0 (no); 1 (yes),
- Month: month indicator 0, 3, 6, 9, 13, 16 and 20 (i.e. factor).
- MonthC: month since initiation of treatment 0, 3, 6, 9, 13, 16 and 20 (i.e. numeric).

Study the output of Model A and Model B and answer the following questions:

a1 What is the difference between Model A and Model B in terms of the correlation structure assumed?

*Answer:*

a2 Which test(s) can be used to test which of these two models fits the data best? Based on the output of these two models, compute the test statistic(s) and specify the asymptotic distribution(s) under the null hypothesis including, if relevant, degrees of freedom.

*Answer:*

a3 Which test(s) can be used for the null hypothesis that the mean retinal sensitivity profiles are the same between group A and group B? Give the name(s) of the test(s) and the asymptotic distribution(s) (with number of degrees of freedom, if relevant) under the null hypothesis.

*Answer:*

A linear mixed effects model has been fitted on the same data assuming for the fixed effects part, linear evolutions in time that differ in the two groups and a random intercepts term to model the within patient correlations.

Study the output in Appendix II of Model C and reply to the following questions:

b1 Give the expressions for the linear mixed effects model. Carefully state all the model assumptions. Introduce and explain your own notation.

*Answer:*

b2 What is the estimated variance of Y at Month 9?

*Answer:*

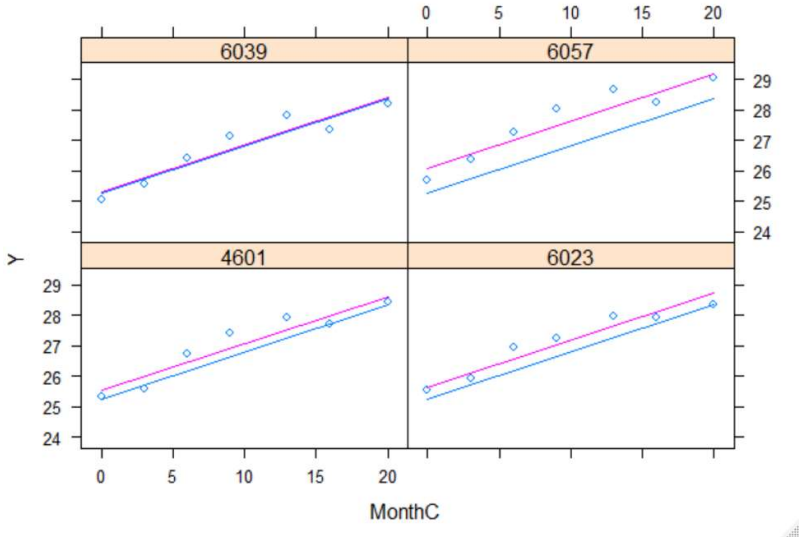
b3 What is the estimated correlation between Y on Month 0 and Y on Month 9?

*Answer:*

b4 Let $y_{ij}$ be retinal sensitivity of patient i (i = 1,…, 20) at month j (j = 0, …, 20) and $b_i$ be the random intercepts term. What is the estimate of $var(y_{i9}|b_i)$?

*Answer:*

c From the fitted model (Model C) two types of predictions can be derived: the marginal and the individualized predictions. See for the relevant output Appendix II. These predictions are shown in the figure below, where the blue lines correspond to the marginal predictions and the pink lines correspond to the individualized predictions. Based on the output of this model compute for patient "6057" his predicted values both marginal and individualized at month 9. Note that patient "6057" was assigned to treatment group A.




*Answer:*

**Make either d1 or d2, but not both.**

d1 To fill in the missing values for the 2 patients, the unconditional mean imputation has been used. Discuss the consequences that this has on the inference (in terms of bias in point estimates and standard errors) using Model C above.

*Answer:*

d2 To fill in the missing values for the 2 patients, the conditional mean imputation has been used. Discuss the consequences that this has on the inference using Model C above.

*Answer:*

..... **from the next four questions, answer one from e1, e2, one from f1, f2, not more** .....

The researchers proceeded further with the analysis of the binary outcome i.e. the presence of retinal sensitivity. They have used the GEE approach, where:

- for the mean part they assumed linear evolutions in time that differ in the two groups and
- for the correlation they used an AR1 correlation matrix.

Study the output of the GEE-1 in Appendix III and answer the following questions:

**Make either e1 or e2, but not both.**

e1 What is the estimated odds ratio for the presence of retinal fluid between group B and A at Months 16? Give an interpretation.

*Answer:*

e2 What is the estimated log-odds ratio for the presence of retinal fluid between group B and A at month 1? Report a 95% confidence interval. Is the log-odds ratio statistically significant at significance level 5%?

*Answer:*

**Make either f1 or f2, but not both.**

f1 The researchers want to test the null hypothesis that the log odds profiles are the same for the two treatments using the multivariate Wald test.

(i) Give the form of the contrast matrix needed to test this null hypothesis.
(ii) What is the asymptotic distribution of the Wald statistic under the null hypothesis in this case?

*Answer:*

f2 The data are analysed again using the GEE approach with the same mean structure as in GEE-1 (i.e. linear evolutions in time that differ in the two groups), but the exchangeable correlation matrix has been used to capture the within patient correlation. The output is given in Appendix III in the part GEE-2. Which of the two GEEs would you prefer for the specific dataset analysed here? Motivate your answer.

*Answer:*

g Is the inference derived with the GEE approach valid, given that two patients have missing values? Explain why.

*Answer:*

The researchers analysed the same binary outcome, i.e. the presence of retinal sensitivity, using the mixed effects logistic regression where the fixed effects part is the same as the mean part of the GEE approach.

Study the output of Model C in Appendix III and answer the following questions:

h Give the expressions for the mixed effects logistic regression model. Carefully state all the model assumptions. Introduce and explain your own notation.

*Answer:*

**Make either i1 or i2, but not both.**

i1 What is the interpretation of the coefficient of the term "`MonthC`"?

*Answer:*

i2 What is the interpretation of the coefficient of the term "`MonthC`" in the output with the marginal coefficients?

*Answer:*

Model C, as presented in Appendix III, is now extended into model D, where a different random effects structure is used. The fixed effects part remains unchanged. Study the output of Model D in Appendix IV and answer the following questions:

j State the assumptions for the random effects part of model D.

*Answer:*

k To test if Model C is equivalent to Model D, state the null and alternative hypothesis you need to test using the notation you introduced in questions h and j. Which test(s) can be used in this case? Give the asymptotic null distribution including degrees of freedom, if relevant.

*Answer:*

**Output for Exercise 3:**

**Long-term follow-up of chronic central serous chorioretinopathy after successful treatment with photodynamic therapy or micropulse laser**

## Appendix I

## Model A

```
model.1 <- gls(Y ~ Month * Treat,
               data = data.c,
               correlation = corSymm(form = ~ 1 | ID),
               weights = varIdent(form = ~ 1 | Month),
               na.action = na.exclude, method = "REML")
summary(model.1)
```

```
Generalized least squares fit by REML
  Model: Y ~ Month * Treat
  Data: data.c
  AIC BIC logLik
   62 178     11

Correlation Structure: General
 Formula: ~1 | ID
 Parameter estimate(s):
 Correlation:
  1     2     3     4     5     6
2 0.929
3 0.885 0.962
4 0.860 0.938 0.966
5 0.824 0.912 0.955 0.977
6 0.828 0.888 0.951 0.948 0.966
7 0.810 0.891 0.955 0.959 0.976 0.982
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | Month
 Parameter estimates:
   0    3    6    9   13   16   20
1.00 1.18 1.11 1.16 1.35 1.37 1.52

Coefficients:
               Value Std.Error t-value p-value
(Intercept)    25.09    0.1543   162.6  0.0000
Month3          0.31    0.0690     4.6  0.0000
Month6          1.31    0.0797    16.5  0.0000
Month9          2.01    0.0935    21.5  0.0000
Month13         2.60    0.1218    21.4  0.0000
Month16         2.33    0.1232    18.9  0.0000
Month20         2.92    0.1454    20.1  0.0000
TreatB          0.01    0.2440     0.1  0.9596
Month3:TreatB  -0.33    0.1090    -3.0  0.0031
Month6:TreatB  -0.90    0.1260    -7.1  0.0000
Month9:TreatB  -0.93    0.1457    -6.4  0.0000
Month13:TreatB -0.90    0.1896    -4.7  0.0000
Month16:TreatB -0.75    0.1918    -3.9  0.0002
Month20:TreatB -0.90    0.2271    -4.0  0.0001


Residual standard error: 0.535
Degrees of freedom: 132 total; 118 residual
```

## Model B

```
model.2 <- gls(Y ~ Month*Treat,
               data = data.c,
               correlation = corCompSymm(form = ~ MonthC | ID),
               weights = varIdent(form = ~ 1 | Month),
               na.action = na.exclude, method = "REML")
summary(model.2)

Generalized least squares fit by REML
  Model: Y ~ Month * Treat
  Data: data.c
     AIC      BIC  logLik
  86.566 147.5211 -21.283

Correlation Structure: Compound symmetry
 Formula: ~MonthC | ID
 Parameter estimate(s):
      Rho
0.9206797
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | Month
 Parameter estimates:
       0        3        6        9       13       16       20
1.000000 1.125754 1.039353 1.074911 1.249763 1.274157 1.417293

Coefficients:
                  Value  Std.Error   t-value p-value
(Intercept)    25.088140 0.16208949 154.77956  0.0000
Month3          0.314394 0.07146740   4.39913  0.0000
Month6          1.314576 0.06612620  19.87981  0.0000
Month9          2.025350 0.07256211  27.91196  0.0000
Month13         2.622481 0.08780577  29.86684  0.0000
Month16         2.358649 0.09045054  26.07667  0.0000
Month20         2.948194 0.10766037  27.38421  0.0000
TreatB          0.012399 0.25628599   0.04838  0.9615
Month3:TreatB  -0.328705 0.11299988  -2.90889  0.0043
Month6:TreatB  -0.896922 0.10455470  -8.57849  0.0000
Month9:TreatB  -0.948918 0.11048397  -8.58874  0.0000
Month13:TreatB -0.915252 0.13409596  -6.82535  0.0000
Month16:TreatB -0.775272 0.13823667  -5.60829  0.0000
Month20:TreatB -0.926509 0.16526983  -5.60604  0.0000

Standardized residuals:
       Min         Q1        Med         Q3        Max
-2.4043424 -0.6397051  0.1652227  0.7726158  1.5216951

Residual standard error: 0.5614945
Degrees of freedom: 132 total; 118 residual
```

## Appendix II

## Model C

```
library(nlme)
model.2 <- lme(Y ~ MonthC*Treat, random = ~1|ID,
               data = data.c, method = "REML")
summary(model.2)
```

```
Linear mixed-effects model fit by REML
 Data: data.c
  AIC BIC logLik
  191 208  -89.3

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:       0.623    0.358

Fixed effects: Y ~ MonthC * Treat
              Value Std.Error  DF t-value p-value
(Intercept)   25.26    0.1928 110   131.0   0.000
MonthC         0.15    0.0064 110    24.3   0.000
TreatB        -0.27    0.3045  18    -0.9   0.382
MonthC:TreatB -0.04    0.0096 110    -4.3   0.000
 Correlation:
              (Intr) MonthC TreatB
MonthC        -0.282
TreatB        -0.633  0.178
MonthC:TreatB  0.187 -0.663 -0.287

Standardized Within-Group Residuals:
   Min     Q1    Med     Q3    Max
-2.725 -0.642 -0.054  0.633  1.856

Number of Observations: 132
Number of Groups: 20
```

```
ranef(model.2)
     (Intercept)
4601      0.2644
6021      0.3710
6023      0.3769
6025      0.2690
6039      0.0426
6045     -0.1692
6049     -0.5129
6053     -0.9084
6055      0.8175
6057      0.8319
6061     -1.2553
6069     -0.4472
6075      0.0980
6077      0.0174
6081      0.3603
6497      0.8180
6499      0.5801
6525     -0.5023
6527     -0.5873
8533     -0.4644
```

## Appendix III

### GEE-1

```
library(geepack)
data.c.new <- data.c[order(data.c$ID, data.c$MonthC), ]
gee1 <- geeglm(Fluid ~ MonthC * Treat,
                  data = data.c.new,
                  id = ID, family = binomial("logit"),
                  corstr = "ar1")
summary(gee1)

Call:
geeglm(formula = Fluid ~ MonthC * Treat, family = binomial("logit"),
    data = data.c.new, id = ID, corstr = "ar1")

 Coefficients:
             Estimate Std.err   Wald Pr(>|W|)
(Intercept)    0.5156  0.6409   0.65     0.42
MonthC         0.3427  0.0323 112.32   <2e-16 ***
TreatB        -0.8313  1.0143   0.67     0.41
MonthC:TreatB -0.0391  0.0690   0.32     0.57
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation structure = ar1
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.473    0.12
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.435  0.0822
Number of clusters:   20  Maximum cluster size: 7


round(vcov(gee1), 3)
             (Intercept) MonthC TreatB MonthC:TreatB
(Intercept)        0.411 -0.021 -0.411         0.021
MonthC            -0.021  0.001  0.021        -0.001
TreatB            -0.411  0.021  1.029        -0.058
MonthC:TreatB      0.021 -0.001 -0.058         0.005
```

## GEE-2

```
gee2 <- geeglm(Fluid ~ MonthC * Treat,
               data = data.c.new,
               id = ID, family = binomial("logit"),
               corstr = "exchangeable")
summary(gee2)

Call:
geeglm(formula = Fluid ~ MonthC * Treat, family = binomial("logit"),
    data = data.c.new, id = ID, corstr = "exchangeable")

 Coefficients:
             Estimate Std.err  Wald Pr(>|W|)
(Intercept)   0.3759  0.6692  0.32     0.57
MonthC        0.3695  0.0460 64.50    1e-15 ***
TreatB       -0.9608  1.0814  0.79     0.37
MonthC:TreatB -0.0298  0.0854  0.12     0.73
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation structure = exchangeable
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.463  0.0874
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha     0.19   0.065
Number of clusters:   20  Maximum cluster size: 7
```

## Appendix IV

### MODEL C

```
library(GLMMadaptive)
model.1 <- mixed_model(Fluid ~ MonthC * Treat,
                       data = data.c.new,
                       random = ~ 1 | ID,
                       family = binomial())
summary(model.1)

Call:
mixed_model(fixed = Fluid ~ MonthC * Treat, random = ~1 | ID,
    data = data.c.new, family = binomial())

Data Descriptives:
Number of Observations: 132
Number of Groups: 20

Model:
 family: binomial
 link: logit

Fit statistics:
 log.Lik  AIC  BIC
     -26 61.9 66.9

Random effects covariance matrix:
            StdDev
(Intercept)   5.24

Fixed effects:
              Estimate Std.Err z-value p-value
(Intercept)      0.209   0.699  0.2996    0.76
MonthC           1.017   0.408  2.4909    0.01
TreatB          -0.110   0.740 -0.1491    0.88
MonthC:TreatB   -0.039   0.405 -0.0962    0.92

Integration:
method: adaptive Gauss-Hermite quadrature rule
quadrature points: 11

Optimization:
method: hybrid EM and quasi-Newton
converged: TRUE


marg <- marginal_coefs(model.1, std_errors = TRUE)
marg
              Estimate Std.Err z-value p-value
(Intercept)    -0.3850   0.611 -0.6300     0.5
MonthC          0.4318   0.279  1.5499     0.1
TreatB         -0.0956   0.679 -0.1409     0.9
MonthC:TreatB  -0.0113   0.309 -0.0367     1.0
```

**MODEL D**

```
model.2 <- mixed_model(Fluid ~ MonthC * Treat,
                       data = data.c.new,
                       random = ~ MonthC | ID,
                       family = binomial)
summary(model.2)

Call:
mixed_model(fixed = Fluid ~ MonthC * Treat, random = ~MonthC |
    ID, data = data.c.new, family = binomial())

Data Descriptives:
Number of Observations: 132
Number of Groups: 20

Model:
 family: binomial
 link: logit

Fit statistics:
 log.Lik  AIC  BIC
   -22.7 59.4 66.3

Random effects covariance matrix:
             StdDev    Corr
(Intercept)  9.2971
MonthC       0.8246 -0.9941

Fixed effects:
              Estimate Std.Err z-value p-value
(Intercept)     0.1854   0.752   0.246    0.81
MonthC          0.9845   0.389   2.533    0.01
TreatB         -0.0961   0.756  -0.127    0.90
MonthC:TreatB  -0.2111   0.352  -0.600    0.55

Integration:
method: adaptive Gauss-Hermite quadrature rule
quadrature points: 11

Optimization:
method: hybrid EM and quasi-Newton
converged: TRUE
```