# LDA_al3998_hw4

AimingLiu

11/9/2020

**In a randomized, double-blind, parallel-group, multicenter study comparing two oral treatments (denoted A and B) for toe-nail infection (De Backer etal., 1998; also see Lesaffre and Spiessons, 2001), patients were evaluated for the degree of onycholysis (the degree of separation of the nail plate from the nail-bed) at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48 thereafter. The onycholysis outcome variable is binary (none or mild versus moderate or severe). The binary outcome was evaluated on 294 patients comprising a total of 1908 measurements. The main objective of the analyses is to compare the effects of oral treatments A and B on changes in the probability of the binary onycholysis outcome over the duration of the study. The raw data are stored in an external file: toenail.dat Each row of the data set contains the following five variables: ID,Y,Treatment,Month,Visit. The binary onycholysis outcome variable Y is coded 0 = none or mild, 1 = moderate or severe. The categorical variable Treatment is coded 1=oral treatment A, 0=oral treatment B. The variable Month denotes the exact timing of measurements in months. The variable Visit denotes the visit number (visit numbers 1-7 correspond to scheduled visits at 0, 4, 8, 12, 24, 36, and 48 weeks).**

## Question 1

**1. Consider a first order transition model for the log odds of moderate or severe onycholysis. Set up a suitable model assuming linear trends. Use month as the time variable.**

```
## Question 1
# load original data
toenail1 <- fread("toenail.dat") %>%
  mutate(Treatment = as.factor(Treatment))
```

```
# add response at lag 1
toenail2 = toenail1 %>%
  group_by(Subject_ID) %>%
  mutate(Response_1 = lag(Response,1))
```

```
## transition probabilities
tab1 <- table(toenail2$Response,toenail2$Response_1)
tab1 %>%
  knitr::kable()
```

| | 0 | 1 |
|---|---|---|
| 0 | 1203 | 112 |
| 1 | 28 | 271 |

```
round(prop.table(tab1,margin = 1),2) %>%
knitr::kable()
```

| | 0 | 1 |
|---|---|---|
| 0 | 0.91 | 0.09 |
| 1 | 0.09 | 0.91 |

```
## association between treatment and the response
tab2 <- table(toenail2$Treatment,toenail2$Response)
round(prop.table(tab2,margin = 1),2)
```

```
      0    1
  0 0.77 0.23
  1 0.80 0.20
```

```
# association between treatment and response stratified by previous response i.e. response at lag 1
temp <- split(toenail2,toenail2$Response_1)
tab3 <- lapply(temp, function(z){table(z$Response,z$Treatment)})
lapply(tab3, function(z){round(prop.table(z,margin = 1),2)})
```

```
$`0`

      0    1
  0 0.48 0.52
  1 0.68 0.32

$`1`

      0    1
  0 0.50 0.50
  1 0.52 0.48
```

I fit the model like this:

$$g[E(y_{ij}|H_{ij})] = logit[E(Y_{ij}|Y_{ij-1} = y_{ij-1})] = X'_{ij} * \beta + \alpha_1 * y_{ij-1} + \alpha_2 * y_{ij-1} * onycholysis$$

```
# Fit the model
model_lag_1 <- gee(Response~ Treatment*(Month+Response_1), corstr = "independence",
                   family = binomial("logit"), id = Subject_ID, data = toenail2)
```

```
          (Intercept)            Treatment1                 Month
          -2.91754387           -0.58731398           -0.09042707
           Response_1      Treatment1:Month Treatment1:Response_1
           4.20408170           -0.05921630            0.69205101
```

```
toenail_sum = round(summary(model_lag_1)$coeff,2)
toenail_sum %>%
  knitr::kable()
```

| | Estimate | Naive S.E. | Naive z | Robust S.E. | Robust z |
|---|---|---|---|---|---|
| (Intercept) | -2.92 | 0.32 | -9.05 | 0.30 | -9.58 |
| Treatment1 | -0.59 | 0.53 | -1.10 | 0.48 | -1.21 |
| Month | -0.09 | 0.04 | -2.24 | 0.04 | -2.31 |
| Response_1 | 4.20 | 0.31 | 13.40 | 0.33 | 12.57 |
| Treatment1:Month | -0.06 | 0.07 | -0.91 | 0.07 | -0.81 |
| Treatment1:Response_1 | 0.69 | 0.52 | 1.33 | 0.49 | 1.40 |

From the value we get from the model with interaction,we can see that the p-value of term `Treatment:Response` and `Treatment:Month` is greater than 0.05,so we drop the interaction terms and fit the model again.

$$g[E(y_{ij}|H_{ij})] = logit[E(Y_{ij}|Y_{ij-1} = y_{ij-1})] = X'_{ij} * \beta + \alpha_1 * y_{ij-1}$$

```
model_lag_1b <- gee(Response~ Treatment + Month + Response_1, corstr = "independence",
                    family = binomial("logit"), id = Subject_ID, data = toenail2)
```

```
(Intercept)  Treatment1       Month  Response_1
 -3.0094361  -0.3090397  -0.1152287   4.4906918
```

```
toenail_sum_final1 = round(summary(model_lag_1b)$coeff,2)
toenail_sum_final1%>%
   knitr::kable()
```

| | Estimate | Naive S.E. | Naive z | Robust S.E. | Robust z |
|---|---|---|---|---|---|
| (Intercept) | -3.01 | 0.27 | -11.24 | 0.25 | -12.01 |
| Treatment1 | -0.31 | 0.21 | -1.46 | 0.18 | -1.74 |
| Month | -0.12 | 0.03 | -3.73 | 0.03 | -3.41 |
| Response_1 | 4.49 | 0.24 | 18.64 | 0.25 | 18.26 |

So the final model is

$$g[E(y_{ij}|H_{ij})] = \beta_0 + \beta_1 Treatment_i + \beta_2 Month_{ij} + \alpha_1 Response_{ij-1}$$

## 2. Repeat the model using a second order transition model. Is there a justification for a second order transition model?

I fit the model like this:

$$g[E(y_{ij}|H_{ij})] = logit[E(Y_{ij}|Y_{ij-1} = y_{ij-1})] = X'_{ij} * \beta + \alpha_1 * y_{ij-1} + \alpha_2 * y_{ij-1} * onycholysis + \alpha_3 * y_{ij-2} * onycholysis$$

```
# add response at lag 2
toenail3 = toenail2 %>%
  group_by(Subject_ID) %>%
  mutate(Response_2 = lag(Response,2))
```

```
## transition probabilities
tab4 <- table(toenail3$Response,toenail3$Response_2)
tab4 %>%
  knitr::kable()
```

| | 0 | 1 |
|---|---|---|
| 0 | 934 | 189 |
| 1 | 33 | 169 |

```
round(prop.table(tab4,margin = 1),2) %>%
knitr::kable()
```

| | 0 | 1 |
|---|---|---|
| 0 | 0.83 | 0.17 |
| 1 | 0.16 | 0.84 |

```
# association between treatment and response stratified by previous response i.e. response at lag 1
temp1 <- split(toenail3,toenail3$Response_2)
tab5 <- lapply(temp, function(z){table(z$Response,z$Treatment)})
lapply(tab5, function(z){round(prop.table(z,margin = 1),2)})
```

```
$`0`

     0    1
0 0.48 0.52
1 0.68 0.32

$`1`

     0    1
0 0.50 0.50
1 0.52 0.48
```

```
# Fit the second order model
model_lag_2 <- gee(Response ~ Treatment *(Month+Response_1+Response_2),
                   corstr = "independence",family = binomial("logit"), id = Subject_ID, data = toenail3)
```

```
          (Intercept)            Treatment1                 Month
          -3.05941334           -0.43737470           -0.05985574
           Response_1            Response_2      Treatment1:Month
           3.29362134            0.74239666           -0.06984192
Treatment1:Response_1 Treatment1:Response_2
           2.36414282           -1.78090434
```

```
toenail_sum_1 = round(summary(model_lag_2)$coeff,2)
toenail_sum_1%>%
  knitr::kable()
```

| | Estimate | Naive S.E. | Naive z | Robust S.E. | Robust z |
|---|---|---|---|---|---|
| (Intercept) | -3.06 | 0.40 | -7.70 | 0.39 | -7.88 |
| Treatment1 | -0.44 | 0.65 | -0.67 | 0.60 | -0.73 |
| Month | -0.06 | 0.04 | -1.35 | 0.04 | -1.34 |
| Response_1 | 3.29 | 0.48 | 6.89 | 0.38 | 8.74 |
| Response_2 | 0.74 | 0.47 | 1.57 | 0.37 | 2.01 |
| Treatment1:Month | -0.07 | 0.07 | -0.95 | 0.08 | -0.88 |
| Treatment1:Response_1 | 2.36 | 1.06 | 2.22 | 0.94 | 2.51 |
| Treatment1:Response_2 | -1.78 | 1.02 | -1.74 | 0.88 | -2.03 |

From the result above, we can get the p-values from the z-value,we can see that the p-value of the interaction term between `Treatment` and `Month` is greater than 0.05, so that we drop this interaction term.And I refit the model as below.

```
model_lag_2b <- gee(Response ~ Month+Treatment*(Response_1+Response_2), corstr = "independence",
                    family = binomial("logit"), id = Subject_ID, data = toenail3)
```

```
          (Intercept)                 Month              Treatment1
          -2.90028180           -0.08588126           -0.83861051
           Response_1            Response_2 Treatment1:Response_1
           3.26711040            0.73980008            2.33507404
Treatment1:Response_2
          -1.68000316
```

```
toenail_sum_2 = round(summary(model_lag_2b)$coeff,2)
toenail_sum_2 %>%
  knitr::kable()
```

| | Estimate | Naive S.E. | Naive z | Robust S.E. | Robust z |
|---|---|---|---|---|---|
| (Intercept) | -2.90 | 0.34 | -8.41 | 0.33 | -8.68 |
| Month | -0.09 | 0.03 | -2.49 | 0.04 | -2.33 |
| Treatment1 | -0.84 | 0.49 | -1.70 | 0.46 | -1.84 |
| Response_1 | 3.27 | 0.47 | 6.96 | 0.38 | 8.71 |
| Response_2 | 0.74 | 0.47 | 1.58 | 0.37 | 2.00 |

| | Estimate | Naive S.E. | Naive z | Robust S.E. | Robust z |
|---|---|---|---|---|---|
| Treatment1:Response_1 | 2.34 | 1.03 | 2.27 | 0.92 | 2.53 |
| Treatment1:Response_2 | -1.68 | 0.99 | -1.70 | 0.85 | -1.98 |

All the term is significant. So the model is :

$$g[E(y_{ij}|H_{ij})] = \beta_0 + \beta_1 Treatment_i + \beta_2 Month_{ij} + \alpha_1 * Response_{ij-1} + \alpha_2 * Response_{ij-2} \\ +\alpha_3 * Treatment_i * Response_{ij-1} + \alpha_4 * Treatment_i * Response_{ij-2}$$

## 3.Provide Interpretations for the parameters in your model.

## For model in Question 1:

$\beta_0$: The log odds of onycholysis outcome at month 0 is `-3.01` for those who receive the oral treatment B and did not have moderate or severe onycholysis outcome at previous visit.

$\beta_1$:The log odds ratio of the onycholysis outcome comparing those receive oral treatment A with oral treatment B who had the idential status for onycholysis outcome in the previous visit with month fixed is `-0.31`.

$\beta_2$: The log odds ratio of the onycholysis outcome with one unit increase in month for those who had the same treatment and the idential status for onycholysis outcome in the previous visit is `-0.12`.

$\alpha_1$ : The log odds ratio of the onycholysis outcome comparing those who had moderate or severe onycholysis with those who had none or mild onycholysis in their previous visit who currently have identical treatment with month fixed is 4.49.

## For model in Question 2:

$\beta_0$: The log odds of onycholysis outcome at month 0 is `-2.9` for those who receive the oral treatment B and have none or mild onycholysis outcome at previous visit and previous two visits.

$\beta_1$:The log odds ratio of the onycholysis outcome comparing those receive oral treatment A with oral treatment B who had none or mild status for onycholysis outcome in the previous visit and previous 2 visits with month fixed is `-0.84`.

$\beta_2$: The log odds ratio of the onycholysis outcome with one unit increase in month for those who had identical treatment and idential status for onycholysis outcome in the previous visit and previous two visits is `-0.09`.

$\alpha_1$: The log odds ratio of the onycholysis outcome comparing those who had moderate or severe onycholysis with those who had none or mild onycholysis in treatment A in their previous visit and who have identical treatment and idential status for onycholysis outcome in the previous two visits with month fixed is `3.27`.

$\alpha_2$: The log odds ratio of the onycholysis outcome comparing those who had moderate or severe onycholysis with those who had none or mild onycholysis in treatment A in their previous two visitw and who have identical treatment and idential status for onycholysis outcome in the previous visit with month fixed is `0.74`.

$\alpha_3$:The difference in log odds ratio of the onycholysis outcome comparing oral treatment A with oral treatment B between a patient who had none or mild response with a patient who had moderate or severe reponse in the previous visit and who have identical treatment and idential status for onycholysis outcome in the previous two visit with month fixed is `2.34`.

$\alpha_4$:The difference in log odds ratio of the onycholysis outcome comparing oral treatment A with oral treatment B between a patient who had none or mild response with a patient who had moderate or severe reponse in the previous two visit and who have identical treatment and idential status for onycholysis outcome in the previous visits with month fixed is `-1.68`.

### 4. How are the interpretations different from the models in HW2 and HW3.

The transition model that is built in this homework focuses more on the effect of the response history of subjects but the random effects model with a random intercept that was built in HW2 and GEE model that was built in HW3 didn't consider the impact of response history.

## Question 2

### 1. Perform a complete case analysis considering a GEE model for the log odds of moderate or severe onycholysis. Set up a suitable model assuming linear trends. Use visit as the time variable.

```
library(tidyr)
library(data.table)
toenail <- fread("toenail.txt")
colnames(toenail) <- c("id","response","treatment","month","visit")
toenail4 <- tidyr::complete(toenail, id, visit) %>%
tidyr::fill(treatment)
toenail4 <- as.data.table(toenail4)
toenail4 %>%
  mutate(response = as.factor(response),
         treatment = as.factor(treatment))
```

```
# center visit
toenail$visit <- toenail$visit - min(toenail$visit)
table(toenail$response,useNA = "always")
```

```
   0    1 <NA>
1500  408    0
```

```
table(toenail$visit,toenail$response,useNA = "always")
```

```
        0   1 <NA>
  0    185 109    0
  1    191  97    0
  2    199  84    0
  3    214  58    0
  4    241  22    0
  5    226  18    0
  6    244  20    0
  <NA>   0   0    0
```

```
# complete case analysis
count <- toenail[,j = list(n=sum(!is.na(response))), by = "id"]
table(count$n)
```

```
  1   2   3   4   5   6   7
  5   3   7   6  10  39 224
```

```
count <- count[n==7]
toenail5<- toenail4[id %in% count$id]
table(toenail5$response,useNA = "always")
```

```
   0    1 <NA>
1266  302    0
```

```
# complete case model
gee1 <- geeglm(response~ treatment* visit , id = id, data = toenail5, family = binomial(link = "logit"))
gee_sum_1 = round(summary(gee1)$coeff,2)
gee_sum_1%>%
  knitr::kable()
```

| | Estimate | Std.err | Wald | Pr(>\|W\|) |
|---|---|---|---|---|
| (Intercept) | 1.056967e+15 | 1.429482e+15 | 0.55 | 0.46 |
| treatment | -7.939057e+15 | 1.905017e+15 | 17.37 | 0.00 |
| visit | -2.835953e+14 | 2.722558e+14 | 1.09 | 0.30 |
| treatment:visit | 1.283241e+15 | 3.740368e+14 | 11.77 | 0.00 |

**2. Perform an available case analysis considering a GEE model for the log odds of moderate or severe onycholysis. Set up a suitable model assuming linear trends. Use visit as the time variable.**

```
toenail6 <- toenail4
table(toenail6$response,useNA = "always")
```

```
   0    1 <NA>
1500  408  150
```

```
table(toenail6$visit,toenail6$response,useNA = "always")
```

```
        0   1 <NA>
  1   185 109    0
  2   191  97    6
  3   199  84   11
  4   214  58   22
  5   241  22   31
  6   226  18   50
  7   244  20   30
  <NA>   0   0    0
```

```r
# available case analysis
gee2 <- geeglm(response ~ treatment * visit, id = id, data = toenail6, family = binomial(link = "logit"
gee_sum_2 =summary(gee2)
gee_sum_2$coefficients %>%
 knitr::kable()
```

| | Estimate | Std.err | Wald | Pr(>\|W\|) |
|---|---|---|---|---|
| (Intercept) | 6.856100e-03 | 7.610757e-01 | 0.0000812 | 0.9928124 |
| treatment | 1.931862e+15 | 4.903722e+14 | 15.5203092 | 0.0000816 |
| visit | -3.494336e-01 | 1.770541e-01 | 3.8950905 | 0.0484274 |
| treatment:visit | -5.733938e+14 | 9.129391e+13 | 39.4477663 | 0.0000000 |

### 3. Perform an LOCF analysis considering a GEE model for the log odds of moderate or severe onycholysis. Set up a suitable model assuming linear trends. Use visit as the time variable.

```r
# LOCF
toenail7 <- lapply(unique(toenail4$id), function(z){tidyr::fill(toenail4[id == z], response)})
toenail7 <- rbindlist(toenail7)
table(toenail7$visit, toenail7$response,useNA = "always")
```

```
        0   1 <NA>
  1   185 109    0
  2   195  99    0
  3   207  87    0
  4   228  66    0
  5   261  33    0
  6   269  25    0
  7   269  25    0
  <NA>  0   0    0
```

```r
#LOCF
gee3 <- geeglm(response~ treatment *  visit , id = id, data = toenail7, family = binomial(link = "logit
gee_sum_3 = round(summary(gee3)$coeff,2)
gee_sum_3%>%
  knitr::kable()
```

| | Estimate | Std.err | Wald | Pr(>\|W\|) |
|---|---|---|---|---|
| (Intercept) | -0.53 | 0.31 | 2.87 | 0.09 |
| treatment | 0.25 | 0.41 | 0.36 | 0.55 |
| visit | -0.21 | 0.05 | 16.99 | 0.00 |
| treatment:visit | -0.12 | 0.07 | 2.60 | 0.11 |

The p-value of the interaction term between visit and treatment is 0.11, which is larger than 0.05. We failed to reject the null hypothesis and conclude that this term should be dropped in the model. The final mixed effects model using multiple imputation should be as the following one.

```
gee3b <- geeglm(response~ treatment +  visit , id = id, data = toenail7, family = binomial(link = "logi
gee_sum_3b = round(summary(gee3b)$coeff,2)
gee_sum_3b%>%
  knitr::kable()
```

| | Estimate | Std.err | Wald | Pr(>\|W\|) |
|---|---|---|---|---|
| (Intercept) | -0.22 | 0.19 | 1.32 | 0.25 |
| treatment | -0.16 | 0.24 | 0.45 | 0.50 |
| visit | -0.29 | 0.03 | 85.17 | 0.00 |

**4. Perform an multiple imputation based analysis considering a GEE model for the log odds of moderate or severe onycholysis. Set up a suitable model assuming linear trends. Use visit as the time variable.**

```
toenail8 <- toenail4
pred <- make.predictorMatrix(toenail8)
pred
```

```
          id visit response treatment month
id         0     1        1         1     1
visit      1     0        1         1     1
response   1     1        0         1     1
treatment  1     1        1         0     1
month      1     1        1         1     0
```

```
pred["response","id"] <- -2
pred
```

```
          id visit response treatment month
id         0     1        1         1     1
visit      1     0        1         1     1
response  -2     1        0         1     1
treatment  1     1        1         0     1
month      1     1        1         1     0
```

```
pred <- pred["response",,drop = FALSE]
pred
```

```
         id visit response treatment month
response -2     1        0         1     1
```

```
toenail8$id <- as.integer(toenail8$id)
imp <- mice(toenail8, method = "2l.bin", pred = pred, seed = 12102, maxit = 1, m = 5, print = FALSE, bl
table(mice::complete(imp)$response, useNA = "always")
```

```
   0    1 <NA>
1500  408  150
```

```
### GEE
implist <- mids2mitml.list(imp)
gee4 <- with(implist, geeglm(response ~ treatment * visit, id=id,family = binomial(link = "logit"), cor
gee4_sum = testEstimates(gee4)
gee4_sum$estimates %>%
  knitr::kable()
```

| | Estimate | Std.Error | t.value | df | P(>\|t\|) | RIV | FMI |
|---|---|---|---|---|---|---|---|
| (Intercept) | 6.856100e-03 | 7.610757e-01 | 0.0090084 | Inf | 0.9928124 | 0 | 0 |
| treatment | 1.931862e+15 | 4.903722e+14 | 3.9395824 | Inf | 0.0000816 | 0 | 0 |
| visit | -3.494336e-01 | 1.770541e-01 | -1.9735984 | Inf | 0.0484274 | 0 | 0 |
| treatment:visit | -5.733938e+14 | 9.129391e+13 | -6.2807457 | Inf | 0.0000000 | 0 | 0 |

**5. Perform an multiple imputation based analysis considering a mixed effects model for the log odds of moderate or severe onycholysis. Set up a suitable model assuming linear trends. Use visit as the time variable.**

```
### Mixed Effects
lme1 <- mice::complete(imp, "all") %>%
purrr::map(lme4::glmer,formula = response ~ treatment * visit + (1 | id),family = binomial) %>%
    pool() %>%
    summary()
lme1 %>%
  knitr::kable()
```

| term | estimate | std.error | statistic | df | p.value |
|---|---|---|---|---|---|
| (Intercept) | -1.4251853 | 0.8659763 | -1.645756 | 1900.804 | 0.0999794 |
| treatment | -0.0101185 | 0.7837087 | -0.012911 | 1900.804 | 0.9897002 |
| visit | -0.8044396 | 0.0892387 | -9.014465 | 1900.804 | 0.0000000 |
| treatment:visit | -0.2351219 | 0.1261251 | -1.864196 | 1900.804 | 0.0624483 |

The p-value of the interaction term between visit and treatment is 0.062, which is larger than 0.05. We failed to reject the null hypothesis and conclude that this term should be dropped in the model. The final mixed effects model using multiple imputation should be as the following one.

```
lme1b <- mice::complete(imp, "all") %>%
purrr::map(lme4::glmer,formula = response ~ treatment +  visit + (1 | id),family = binomial) %>%
    pool() %>%
    summary()
lme1b %>%
  knitr::kable()
```

| term | estimate | std.error | statistic | df | p.value |
|---|---|---|---|---|---|
| (Intercept) | -1.0518224 | 0.7993851 | -1.315789 | 1901.804 | 0.1884033 |
| treatment | -0.6968836 | 0.6869498 | -1.014461 | 1901.804 | 0.3104920 |
| visit | -0.9115252 | 0.0743365 | -12.262151 | 1901.804 | 0.0000000 |

## Appendix: code

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE, comment = "")
library(tidyverse)
library(data.table)
library(geepack)
library(gee)
library(ipw)
library(doBy)
library(mice)
library(purrr)
library(mitml)
library(CRTgeeDR)
library(broom.mixed)
## Question 1
# load original data
toenail1 <- fread("toenail.dat") %>%
  mutate(Treatment = as.factor(Treatment))
# add response at lag 1
toenail2 = toenail1 %>%
  group_by(Subject_ID) %>%
  mutate(Response_1 = lag(Response,1))
## transition probabilities
tab1 <- table(toenail2$Response,toenail2$Response_1)
tab1 %>%
  knitr::kable()
round(prop.table(tab1,margin = 1),2) %>%
knitr::kable()
## association between treatment and the response
tab2 <- table(toenail2$Treatment,toenail2$Response)
round(prop.table(tab2,margin = 1),2)
# association between treatment and response stratified by previous response i.e. response at lag 1
temp <- split(toenail2,toenail2$Response_1)
tab3 <- lapply(temp, function(z){table(z$Response,z$Treatment)})
lapply(tab3, function(z){round(prop.table(z,margin = 1),2)})
# Fit the model
model_lag_1 <- gee(Response~ Treatment*(Month+Response_1), corstr = "independence",
                   family = binomial("logit"), id = Subject_ID, data = toenail2)
toenail_sum = round(summary(model_lag_1)$coeff,2)
toenail_sum %>%
  knitr::kable()
model_lag_1b <- gee(Response~ Treatment + Month + Response_1, corstr = "independence",
                    family = binomial("logit"), id = Subject_ID, data = toenail2)
toenail_sum_final1 = round(summary(model_lag_1b)$coeff,2)
toenail_sum_final1%>%
   knitr::kable()
# add response at lag 2
toenail3 = toenail2 %>%
  group_by(Subject_ID) %>%
  mutate(Response_2 = lag(Response,2))
## transition probabilities
tab4 <- table(toenail3$Response,toenail3$Response_2)
tab4 %>%
```

```
  knitr::kable()
round(prop.table(tab4,margin = 1),2) %>%
knitr::kable()
# association between treatment and response stratified by previous response i.e. response at lag 1
temp1 <- split(toenail3,toenail3$Response_2)
tab5 <- lapply(temp, function(z){table(z$Response,z$Treatment)})
lapply(tab5, function(z){round(prop.table(z,margin = 1),2)})
# Fit the second order model
model_lag_2 <- gee(Response ~ Treatment *(Month+Response_1+Response_2),
                   corstr = "independence",family = binomial("logit"), id = Subject_ID, data = toenail3)
toenail_sum_1 = round(summary(model_lag_2)$coeff,2)
toenail_sum_1%>%
  knitr::kable()
model_lag_2b <- gee(Response ~ Month+Treatment*(Response_1+Response_2), corstr = "independence",
                    family = binomial("logit"), id = Subject_ID, data = toenail3)
toenail_sum_2 = round(summary(model_lag_2b)$coeff,2)
toenail_sum_2 %>%
  knitr::kable()
library(tidyr)
library(data.table)
toenail <- fread("toenail.txt")
colnames(toenail) <- c("id","response","treatment","month","visit")
toenail4 <- tidyr::complete(toenail, id, visit) %>%
tidyr::fill(treatment)
toenail4 <- as.data.table(toenail4)
toenail4 %>%
  mutate(response = as.factor(response),
         treatment = as.factor(treatment))
# center visit
toenail$visit <- toenail$visit - min(toenail$visit)
table(toenail$response,useNA = "always")
table(toenail$visit,toenail$response,useNA = "always")
# complete case analysis
count <- toenail[,j = list(n=sum(!is.na(response))), by = "id"]
table(count$n)
count <- count[n==7]
toenail5<- toenail4[id %in% count$id]
table(toenail5$response,useNA = "always")
# complete case model
gee1 <- geeglm(response~ treatment* visit , id = id, data = toenail5, family = binomial(link = "logit")
gee_sum_1 = round(summary(gee1)$coeff,2)
gee_sum_1%>%
  knitr::kable()
toenail6 <- toenail4
table(toenail6$response,useNA = "always")
table(toenail6$visit,toenail6$response,useNA = "always")
# available case analysis
gee2 <- geeglm(response ~ treatment * visit, id = id, data = toenail6, family = binomial(link = "logit"
gee_sum_2 =summary(gee2)
gee_sum_2$coefficients %>%
 knitr::kable()
# LOCF
toenail7 <- lapply(unique(toenail4$id), function(z){tidyr::fill(toenail4[id == z], response)})
```

```r
toenail7 <- rbindlist(toenail7)
table(toenail7$visit, toenail7$response,useNA = "always")
#LOCF
gee3 <- geeglm(response~ treatment *  visit , id = id, data = toenail7, family = binomial(link = "logit
gee_sum_3 = round(summary(gee3)$coeff,2)
gee_sum_3%>%
  knitr::kable()
gee3b <- geeglm(response~ treatment +  visit , id = id, data = toenail7, family = binomial(link = "logi
gee_sum_3b = round(summary(gee3b)$coeff,2)
gee_sum_3b%>%
  knitr::kable()
toenail8 <- toenail4
pred <- make.predictorMatrix(toenail8)
pred
pred["response","id"] <- -2
pred
pred <- pred["response",,drop = FALSE]
pred
toenail8$id <- as.integer(toenail8$id)
imp <- mice(toenail8, method = "2l.bin", pred = pred, seed = 12102, maxit = 1, m = 5, print = FALSE, bl
table(mice::complete(imp)$response, useNA = "always")
### GEE
implist <- mids2mitml.list(imp)
gee4 <- with(implist, geeglm(response ~ treatment * visit, id=id,family = binomial(link = "logit"), cor
gee4_sum = testEstimates(gee4)
gee4_sum$estimates %>%
  knitr::kable()
### Mixed Effects
lme1 <- mice::complete(imp, "all") %>%
purrr::map(lme4::glmer,formula = response ~ treatment * visit + (1 | id),family = binomial) %>%
    pool() %>%
    summary()
lme1 %>%
  knitr::kable()
lme1b <- mice::complete(imp, "all") %>%
purrr::map(lme4::glmer,formula = response ~ treatment +  visit + (1 | id),family = binomial) %>%
    pool() %>%
    summary()
lme1b %>%
  knitr::kable()
```