Analysis

Preliminaries

Import libraries to import, manipulate and analyze the dataset

```
> library("readr")
> options(tidyverse.quiet = TRUE)
> library("tidyverse")
> ## package testthat needed to check whether two R objects are
> ## equal/identical
> library("testthat", warn.conflicts = FALSE)
> library("multgee", quietly = TRUE)
> library("VGAM", quietly = TRUE)
```

Import the dataset and create unique identification variable by combining variables `id` and `vertebra`

```
> lesions <- read_csv("xray_spa_dc_synd_Osteo_vert3s.csv", show_col_types = FALSE)
> lesions <- lesions |>
+     group_by(id, vertebra) |>
+     mutate(id_vertebra = cur_group_id()) |>
+     ungroup()
```

Descriptive Statistics

As a sanity check, tried to replicate the descriptive statistics analysis

```
> lesions |>
+   _$Osteo_synm_tot_corr_upp_low_2 |>
+   table(useNA = "always")

   0    1    2 <NA>
9393  847  280    0
> lesions |>
+   _$Osteo_synm_tot_corr_upp_low_2 |>
+   table() |>
+   prop.table()

        0          1          2
0.89287072 0.08051331 0.02661597
> lesions |>
+   filter(synm_tot_lag == 1) |>
+   _$Osteo_synm_tot_corr_upp_low_2 |>
+   table(useNA = "always")

   0    1    2 <NA>
  38    5   49    0
> lesions |>
+   filter(synm_tot_lag == 1) |>
```

```
+     _$Osteo_synm_tot_corr_upp_low_2 |>
+     table() |>
+     prop.table()

          0          1          2
0.41304348 0.05434783 0.53260870
> lesions |>
+     with(table(synm_tot_lag, Osteo_synm_tot_corr_upp_low_2)) |>
+     print()
              Osteo_synm_tot_corr_upp_low_2
synm_tot_lag     0     1     2
           0  5772   658   152
           1    38     5    49
> lesions |>
+     filter(!is.na(synm_tot_lag)) |>
+     _$id |>
+     n_distinct()
[1] 324
> ## This is different from the reported number of 3696
> lesions |>
+     _$id_vertebra |>
+     n_distinct()
[1] 3846
```

The only different number is the last one, suggesting that the number of unique values `id_vertebra` in the final dataset are 3846 and not 3696. This discrepancy might have occurred due to the way `id_vertebra` was initially constructed.

Finally, we drop variables from the dataset that are not needed

```
> lesions <- lesions |>
+     select(-id, -bdmard, -synm_tot, -vertebra)
```

Analysis using `multgee`

The function `nomLORgee` treats the last category as baseline/reference category. Hence, in the previous code `Osteo_synm_tot_corr_upp_low_2 == 2` was treated as the baseline category and the logits that were created were the following two: i) 0 vs 2 and ii) 1 vs 2. Although this does not affect the inference (i.e. the conclusions are identical), the following code will force `Osteo_synm_tot_corr_upp_low_2 == 0` to be the baseline category and the logits will be: i) 1 vs 0 and ii) 2 vs 0 as in `SAS`:

```
> lesions <-
+     lesions |>
+     mutate(Osteo_synm_tot_corr_upp_low_2 =
+               factor(Osteo_synm_tot_corr_upp_low_2, levels = c("1", "2", "0")),
+           Osteo_synm_tot_corr_upp_low_2_lag =
+               factor(Osteo_synm_tot_corr_upp_low_2_lag),
+           synm_tot_lag =
+               factor(synm_tot_lag),
+           sexe =
+               factor(sexe),
+           hla =
+               factor(hla),
+           tabac_10y =
```

```
+                factor(tabac_10y),
+            profession =
+                factor(profession),
+            bdmard_lag =
+                factor(bdmard_lag))
```

Univariate Analysis

Fitting all univariable models

```
> fit_null <- nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ 1, data = lesions,
+     id = id_vertebra, repeated = t_new, LORstr = "time.exch")
>
> fit_null |>
+     update(formula = Osteo_synm_tot_corr_upp_low_2 ~ synm_tot_lag) |>
+     summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10


Link : Baseline Category Logit

Local Odds Ratios:
Structure:        time.exch
Model:            2way
Homogenous scores: TRUE


call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ synm_tot_lag,
    data = lesions, id = id_vertebra, repeated = t_new, LORstr = "time.exch")


Summary of residuals:
       Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -0.5107842  -0.1006521  -0.0232656  -0.0002243  -0.0232656   0.9767344


Number of Iterations: 4


Coefficients:
                 Estimate    san.se      san.z Pr(>|san.z|)
beta10           -2.16379   0.05142   -42.0826       <2e-16 ***
synm_tot_lag1:1  -0.06467   0.53319    -0.1213       0.9035
beta20           -3.62849   0.09808   -36.9936       <2e-16 ***
synm_tot_lag1:2   3.77391   0.24889    15.1628       <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Local Odds Ratios Estimates:
      [,1]   [,2]  [,3]   [,4]
[1,] 0.000  0.000 1.331  2.671
[2,] 0.000  0.000 2.671 29.269
[3,] 1.331  2.671 0.000  0.000
[4,] 2.671 29.269 0.000  0.000


p-value of Null model: < 0.0001
> fit_null |>
```

```
+       update(Osteo_synm_tot_corr_upp_low_2 ~ Osteo_synm_tot_corr_upp_low_2_lag) |>
+       summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10


Link : Baseline Category Logit

Local Odds Ratios:
Structure:         time.exch
Model:             2way
Homogenous scores: TRUE


call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ Osteo_synm_tot_corr_upp_low_2_lag,
    data = lesions, id = id_vertebra, repeated = t_new, LORstr = "time.exch")


Summary of residuals:
     Min.    1st Qu.    Median       Mean    3rd Qu.        Max.
-0.819581  -0.061935  -0.014864  -0.003271  -0.014864   0.985136


Number of Iterations: 14


Coefficients:
                                  Estimate    san.se     san.z Pr(>|san.z|)
beta10                            -2.70176   0.06213  -43.4881      < 2e-16
Osteo_synm_tot_corr_upp_low_2_lag1:1   4.04996   0.15178   26.6822      < 2e-16
Osteo_synm_tot_corr_upp_low_2_lag2:1   1.10667   0.32054    3.4525      0.00056
beta20                            -4.12888   0.11965  -34.5081      < 2e-16
Osteo_synm_tot_corr_upp_low_2_lag1:2   1.95537   0.40634    4.8122      < 2e-16
Osteo_synm_tot_corr_upp_low_2_lag2:2   5.82712   0.28883   20.1752      < 2e-16

beta10                                 ***
Osteo_synm_tot_corr_upp_low_2_lag1:1 ***
Osteo_synm_tot_corr_upp_low_2_lag2:1 ***
beta20                                 ***
Osteo_synm_tot_corr_upp_low_2_lag1:2 ***
Osteo_synm_tot_corr_upp_low_2_lag2:2 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
          [,1]     [,2]        [,3]      [,4]
[1,]         0    0.000 3505970777     0.000
[2,]         0    0.000          0   240.078
[3,] 3505970777   0.000          0     0.000
[4,]         0  240.078          0     0.000


p-value of Null model: < 0.0001
> fit_null |>
+       update(formula = Osteo_synm_tot_corr_upp_low_2 ~ sexe) |>
+       summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10
```

```
Link : Baseline Category Logit

Local Odds Ratios:
Structure:         time.exch
Model:             3way
Homogenous scores: TRUE

call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ sexe, data = lesions,
    id = id_vertebra, repeated = t_new, LORstr = "time.exch")

Summary of residuals:
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.0946008 -0.0658809 -0.0480322 -0.0005866 -0.0087458  0.9912542

Number of Iterations: 2

Coefficients:
         Estimate    san.se    san.z Pr(>|san.z|)
beta10   -2.24900   0.06581 -34.1717      < 2e-16 ***
sexe1:1  -0.34996   0.10391  -3.3681      0.00076 ***
beta20   -4.63009   0.20180 -22.9439      < 2e-16 ***
sexe1:2   1.71515   0.22382   7.6631      < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
       [,1]     [,2]  [,3]     [,4]  [,5]     [,6]
[1,] 0.000    0.000 1.001    0.926 1.001    0.926
[2,] 0.000    0.000 0.926 148.346 0.926 148.346
[3,] 1.001    0.926 0.000    0.000 1.001    0.926
[4,] 0.926 148.346 0.000    0.000 0.926 148.346
[5,] 1.001    0.926 1.001    0.926 0.000    0.000
[6,] 0.926 148.346 0.926 148.346 0.000    0.000

p-value of Null model: < 0.0001
> fit_null |>
+     update(formula = Osteo_synm_tot_corr_upp_low_2 ~ age_m0) |>
+     summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10

Link : Baseline Category Logit

Local Odds Ratios:
Structure:         time.exch
Model:             3way
Homogenous scores: TRUE

call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ age_m0, data = lesions,
    id = id_vertebra, repeated = t_new, LORstr = "time.exch")
```

```
Summary of residuals:
      Min.    1st Qu.     Median        Mean    3rd Qu.       Max.
-0.6689137 -0.0493256 -0.0291397  0.0005234 -0.0177814  0.9910543


Number of Iterations: 3


Coefficients:
          Estimate   san.se     san.z Pr(>|san.z|)
beta10     -6.97892  0.25662 -27.1958    < 2.2e-16 ***
age_m0:1    0.11977  0.00641  18.6986    < 2.2e-16 ***
beta20     -5.70551  0.36249 -15.7397    < 2.2e-16 ***
age_m0:2    0.06274  0.00931   6.7369    < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Local Odds Ratios Estimates:
            [,1]      [,2]        [,3]      [,4]        [,5]      [,6]
[1,]           0    0.000  2746658398    0.000  2746658398    0.000
[2,]           0    0.000           0  340.557           0  340.557
[3,]  2746658398    0.000           0    0.000  2746658398    0.000
[4,]           0  340.557           0    0.000           0  340.557
[5,]  2746658398    0.000  2746658398    0.000           0    0.000
[6,]           0  340.557           0  340.557           0    0.000


p-value of Null model: < 0.0001
> fit_null |>
+     update(formula = Osteo_synm_tot_corr_upp_low_2 ~ bmi) |>
+     summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10


Link : Baseline Category Logit


Local Odds Ratios:
Structure:         time.exch
Model:             3way
Homogenous scores: TRUE


call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ bmi, data = lesions,
    id = id_vertebra, repeated = t_new, LORstr = "time.exch")


Summary of residuals:
      Min.    1st Qu.     Median        Mean    3rd Qu.       Max.
-0.1571537 -0.0738288 -0.0474298 -0.0002107 -0.0221240  0.9878167


Number of Iterations: 3


Coefficients:
        Estimate   san.se     san.z Pr(>|san.z|)
beta10  -3.93022  0.28203 -13.9354    < 2.2e-16 ***
bmi:1    0.06264  0.01122   5.5834    < 2.2e-16 ***
beta20  -6.16999  0.41339 -14.9254    < 2.2e-16 ***
```

```
bmi:2    0.11054   0.01597    6.9216     < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
            [,1]     [,2]        [,3]      [,4]        [,5]      [,6]
[1,]           0    0.000 2746658914     0.000 2746658914     0.000
[2,]           0    0.000          0   340.557          0   340.557
[3,] 2746658914    0.000          0     0.000 2746658914     0.000
[4,]           0  340.557          0     0.000          0   340.557
[5,] 2746658914    0.000 2746658914     0.000          0     0.000
[6,]           0  340.557          0   340.557          0     0.000


p-value of Null model: < 0.0001
> fit_null |>
+     update(formula = Osteo_synm_tot_corr_upp_low_2 ~ hla) |>
+     summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10

Link : Baseline Category Logit

Local Odds Ratios:
Structure:          time.exch
Model:              3way
Homogenous scores: TRUE

call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ hla, data = lesions,
    id = id_vertebra, repeated = t_new, LORstr = "time.exch")

Summary of residuals:
     Min.    1st Qu.    Median      Mean   3rd Qu.      Max.
-0.114300 -0.063462 -0.029055 -0.000396 -0.025689  0.974311

Number of Iterations: 2

Coefficients:
        Estimate    san.se    san.z Pr(>|san.z|)
beta10 -2.01420   0.07423 -27.1351        <2e-16 ***
hla1:1 -0.64974   0.10242  -6.3438        <2e-16 ***
beta20 -3.38384   0.14722 -22.9847        <2e-16 ***
hla1:2 -0.18448   0.18274  -1.0095        0.3127
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
       [,1]     [,2]  [,3]     [,4]  [,5]     [,6]
[1,] 0.000    0.000 1.001    0.926 1.001    0.926
[2,] 0.000    0.000 0.926  148.346 0.926  148.346
[3,] 1.001    0.926 0.000    0.000 1.001    0.926
[4,] 0.926  148.346 0.000    0.000 0.926  148.346
[5,] 1.001    0.926 1.001    0.926 0.000    0.000
```

```
[6,] 0.926 148.346 0.926 148.346 0.000    0.000

p-value of Null model: < 0.0001
> fit_null |>
+     update(formula = Osteo_synm_tot_corr_upp_low_2 ~ tabac_10y) |>
+     summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10

Link : Baseline Category Logit

Local Odds Ratios:
Structure:         time.exch
Model:             3way
Homogenous scores: TRUE

call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ tabac_10y,
    data = lesions, id = id_vertebra, repeated = t_new, LORstr = "time.exch")

Summary of residuals:
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.0844504 -0.0741461 -0.0374903 -0.0005436 -0.0208306  0.9791694

Number of Iterations: 3

Coefficients:
             Estimate    san.se     san.z Pr(>|san.z|)
beta10       -2.36035   0.06866 -34.3780      < 2e-16 ***
tabac_10y1:1 -0.12300   0.10392  -1.1836      0.23657
beta20       -3.76009   0.12869 -29.2192      < 2e-16 ***
tabac_10y1:2  0.59479   0.17278   3.4424      0.00058 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
           [,1]    [,2]       [,3]    [,4]       [,5]    [,6]
[1,]          0   0.000 2746659173   0.000 2746659173   0.000
[2,]          0   0.000          0 340.557          0 340.557
[3,] 2746659173   0.000          0   0.000 2746659173   0.000
[4,]          0 340.557          0   0.000          0 340.557
[5,] 2746659173   0.000 2746659173   0.000          0   0.000
[6,]          0 340.557          0 340.557          0   0.000

p-value of Null model: 0.00022557
> fit_null |>
+     update(formula = Osteo_synm_tot_corr_upp_low_2 ~ profession) |>
+     summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10

Link : Baseline Category Logit
```

```
Local Odds Ratios:
Structure:          time.exch
Model:              3way
Homogenous scores: TRUE


call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ profession,
    data = lesions, id = id_vertebra, repeated = t_new, LORstr = "time.exch")

Summary of residuals:
      Min.    1st Qu.     Median       Mean     3rd Qu.        Max.
-0.0918982 -0.0918982 -0.0354196 -0.0005317 -0.0245588   0.9754412


Number of Iterations: 3

Coefficients:
              Estimate     san.se      san.z  Pr(>|san.z|)
beta10        -2.76670   0.16734  -16.5332       < 2e-16 ***
profession2:1  0.50344   0.17632    2.8552       0.00430 **
profession3:1 -0.50524   0.28942   -1.7457       0.08086 .
beta20        -3.23084   0.20487  -15.7702       < 2e-16 ***
profession2:2 -0.35203   0.23082   -1.5251       0.12723
profession3:2 -0.17980   0.30807   -0.5836       0.55947
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
       [,1]     [,2]  [,3]     [,4]  [,5]     [,6]
[1,] 0.000    0.000 1.003    1.122 1.003    1.122
[2,] 0.000    0.000 1.122  110.020 1.122  110.020
[3,] 1.003    1.122 0.000    0.000 1.003    1.122
[4,] 1.122  110.020 0.000    0.000 1.122  110.020
[5,] 1.003    1.122 1.003    1.122 0.000    0.000
[6,] 1.122  110.020 1.122  110.020 0.000    0.000


p-value of Null model: < 0.0001
> fit_null |>
+     update(formula = Osteo_synm_tot_corr_upp_low_2 ~ bdmard_lag) |>
+     summary()
Warning in vglm.fitter(x = x, y = y, w = w, offset = offset, Xm2 = Xm2, :
iterations terminated because half-step sizes are very small
Warning in vglm.fitter(x = x, y = y, w = w, offset = offset, Xm2 = Xm2, : some
quantities such as z, residuals, SEs may be inaccurate due to convergence at a
half-step
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10


Link : Baseline Category Logit

Local Odds Ratios:
Structure:          time.exch
Model:              2way
Homogenous scores: TRUE
```

```
call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ bdmard_lag,
    data = lesions, id = id_vertebra, repeated = t_new, LORstr = "time.exch")

Summary of residuals:
     Min.    1st Qu.    Median       Mean   3rd Qu.       Max.
-0.131044 -0.091216 -0.042113 -0.000624 -0.029345  0.970655

Number of Iterations: 3

Coefficients:
               Estimate    san.se    san.z Pr(>|san.z|)
beta10         -2.26606   0.05528 -40.9897      <2e-16 ***
bdmard_lag1:1   0.42397   0.05722   7.4096      <2e-16 ***
beta20         -3.40017   0.09319 -36.4865      <2e-16 ***
bdmard_lag1:2   0.42290   0.10569   4.0015       6e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
          [,1]    [,2]       [,3]    [,4]
[1,]         0   0.000 5652373643   0.000
[2,]         0   0.000          0 260.308
[3,] 5652373643   0.000          0   0.000
[4,]         0 260.308          0   0.000

p-value of Null model: < 0.0001
```

Multivariable Analysis

Fitting the multivariable model

```
> model_gee_multinomial <-
+   fit_null |>
+   update(formula = Osteo_synm_tot_corr_upp_low_2 ~ synm_tot_lag +
+           Osteo_synm_tot_corr_upp_low_2_lag + sexe + age_m0 + bmi + hla +
+           tabac_10y + profession + bdmard_lag)
> model_gee_multinomial |> summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10

Link : Baseline Category Logit

Local Odds Ratios:
Structure:        time.exch
Model:            2way
Homogenous scores: TRUE

call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ synm_tot_lag +
    Osteo_synm_tot_corr_upp_low_2_lag + sexe + age_m0 + bmi +
    hla + tabac_10y + profession + bdmard_lag, data = lesions,
    id = id_vertebra, repeated = t_new, LORstr = "time.exch")
```

```
Summary of residuals:
     Min.   1st Qu.    Median      Mean   3rd Qu.       Max.
-0.990219 -0.044342 -0.020043 -0.003652 -0.006598  0.997553


Number of Iterations: 13

Coefficients:
                                    Estimate  san.se   san.z Pr(>|san.z|)
beta10                              -6.30746 0.48799 -12.9253     < 2e-16
synm_tot_lag1:1                     -0.16356 0.50021  -0.3270     0.74368
Osteo_synm_tot_corr_upp_low_2_lag1:1 3.62272 0.16474  21.9904     < 2e-16
Osteo_synm_tot_corr_upp_low_2_lag2:1 0.89957 0.42053   2.1391     0.03243
sexe1:1                             -0.16897 0.11545  -1.4636     0.14330
age_m0:1                             0.07233 0.00666  10.8631     < 2e-16
bmi:1                                0.03184 0.01420   2.2427     0.02492
hla1:1                              -0.26541 0.11330  -2.3426     0.01915
tabac_10y1:1                         0.03531 0.11067   0.3190     0.74969
profession2:1                        0.35439 0.18131   1.9546     0.05063
profession3:1                        0.21802 0.28469   0.7658     0.44378
bdmard_lag1:1                        0.53880 0.08053   6.6906     < 2e-16
beta20                              -8.87886 0.92830  -9.5646     < 2e-16
synm_tot_lag1:2                      2.09588 0.47003   4.4591       1e-05
Osteo_synm_tot_corr_upp_low_2_lag1:2 2.14007 0.29993   7.1352     < 2e-16
Osteo_synm_tot_corr_upp_low_2_lag2:2 5.50144 0.38498  14.2903     < 2e-16
sexe1:2                              1.63848 0.21124   7.7564     < 2e-16
age_m0:2                             0.06048 0.01494   4.0477       5e-05
bmi:2                                0.03357 0.02743   1.2242     0.22089
hla1:2                              -0.03670 0.21078  -0.1741     0.86178
tabac_10y1:2                        -0.33724 0.19659  -1.7154     0.08626
profession2:2                        0.81317 0.25792   3.1528     0.00162
profession3:2                        1.36909 0.36759   3.7245     0.00020
bdmard_lag1:2                        0.54981 0.17362   3.1668     0.00154

beta10                              ***
synm_tot_lag1:1
Osteo_synm_tot_corr_upp_low_2_lag1:1 ***
Osteo_synm_tot_corr_upp_low_2_lag2:1 *
sexe1:1
age_m0:1                            ***
bmi:1                               *
hla1:1                              *
tabac_10y1:1
profession2:1                        .
profession3:1
bdmard_lag1:1                       ***
beta20                              ***
synm_tot_lag1:2                     ***
Osteo_synm_tot_corr_upp_low_2_lag1:2 ***
Osteo_synm_tot_corr_upp_low_2_lag2:2 ***
sexe1:2                             ***
age_m0:2                            ***
bmi:2
hla1:2
```

```
tabac_10y1:2                                   .
profession2:2                        **
profession3:2                        ***
bdmard_lag1:2                        **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
           [,1]    [,2]        [,3]    [,4]
[1,]          0   0.000 5207136860    0.000
[2,]          0   0.000          0  247.558
[3,] 5207136860   0.000          0    0.000
[4,]          0 247.558          0    0.000


p-value of Null model: < 0.0001
```

and the exponentiated coefficients

```
> model_gee_multinomial |>
+     coefficients() |>
+     exp()
                           beta10                          synm_tot_lag1:1
                     1.822664e-03                             8.491144e-01
Osteo_synm_tot_corr_upp_low_2_lag1:1 Osteo_synm_tot_corr_upp_low_2_lag2:1
                     3.743915e+01                             2.458537e+00
                          sexe1:1                                  age_m0:1
                     8.445366e-01                             1.075008e+00
                            bmi:1                                    hla1:1
                     1.032355e+00                             7.668925e-01
                      tabac_10y1:1                             profession2:1
                     1.035941e+00                             1.425310e+00
                     profession3:1                             bdmard_lag1:1
                     1.243613e+00                             1.713941e+00
                           beta20                          synm_tot_lag1:2
                     1.393026e-04                             8.132622e+00
Osteo_synm_tot_corr_upp_low_2_lag1:2 Osteo_synm_tot_corr_upp_low_2_lag2:2
                     8.500038e+00                             2.450440e+02
                          sexe1:2                                  age_m0:2
                     5.147347e+00                             1.062348e+00
                            bmi:2                                    hla1:2
                     1.034144e+00                             9.639678e-01
                      tabac_10y1:2                             profession2:2
                     7.137399e-01                             2.255052e+00
                     profession3:2                             bdmard_lag1:2
                     3.931778e+00                             1.732929e+00
```

Additional Checks

nomLORgee

We first check the function `nomLORgee`. To test its functionality, we compare the GEE model setting `LORstr == "indepedence"` with the corresponding maximum likelihood (ML) model that treats all clusters (and hence observations as independent). According to the theory that the two models should produce identical estimates for the regression parameters and fitted values.

We fit the GEE model with independence local odds ratios structure

```
> model_gee_independence <- model_gee_multinomial |>
+     update(LORstr = "independence")
```

and the ML model

```
> model_ml <- vglm(formula = Osteo_synm_tot_corr_upp_low_2 ~ synm_tot_lag +
+     Osteo_synm_tot_corr_upp_low_2_lag + sexe + age_m0 + bmi + hla + tabac_10y +
+     profession + bdmard_lag, data = lesions, family = multinomial(ref = "0"))
```

Their coefficients are identical

```
> model_gee_independence_coef <- model_gee_independence |>
+     coefficients() |>
+     matrix(12, 2)
> model_ml_coef <- model_ml |>
+     coefficients() |>
+     matrix(12, 2, TRUE)
> expect_equal(model_gee_independence_coef, model_ml_coef)
```

and so are their fitted values

```
> fitted_values_gee <- model_gee_independence |>
+     fitted.values() |>
+     as.numeric()
> fitted_values_ml <- model_ml |>
+     fitted.values() |>
+     as.numeric()
> expect_equal(fitted_values_gee, fitted_values_ml)
```

The above confirm that the function `nomLORgee` works as expected.

Different reference categories

The output of the model `model_gee_independence` uses as baseline/reference category category 0.

The following code fits a gee model with category 2 as baseline/reference category (as done previously but with `LORstr = "independence"`).

```
> lesions2 <-
+   lesions |>
+   mutate(Osteo_synm_tot_corr_upp_low_2 =
+           factor(Osteo_synm_tot_corr_upp_low_2, levels = c("0", "1", "2")))
> model_gee_multinomial2 <-
+   nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ synm_tot_lag +
+               Osteo_synm_tot_corr_upp_low_2_lag + sexe + age_m0 + bmi + hla +
+               tabac_10y + profession + bdmard_lag,
+            data = lesions2,
+            id = id_vertebra,
+            repeated = t_new,
+            LORstr = "independence")
> model_gee_multinomial2 |> summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10


Link : Baseline Category Logit
```

```
Local Odds Ratios:
Structure:            independence
Homogenous scores: TRUE

call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ synm_tot_lag +
    Osteo_synm_tot_corr_upp_low_2_lag + sexe + age_m0 + bmi +
    hla + tabac_10y + profession + bdmard_lag, data = lesions2,
    id = id_vertebra, repeated = t_new, LORstr = "independence")

Summary of residuals:
    Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
-0.98431  -0.03956  -0.00990   0.00000  0.04791   0.99525

Number of Iterations: 1

Coefficients:
                                          Estimate   san.se    san.z  Pr(>|san.z|)
beta10                                     8.70402   0.98462   8.8400     < 2e-16
synm_tot_lag1:1                           -1.93394   0.53170  -3.6373     0.00028
Osteo_synm_tot_corr_upp_low_2_lag1:1      -1.44341   0.63914  -2.2584     0.02392
Osteo_synm_tot_corr_upp_low_2_lag2:1      -6.12105   0.40832 -14.9909     < 2e-16
sexe1:1                                   -1.59995   0.22586  -7.0840     < 2e-16
age_m0:1                                  -0.05704   0.01480  -3.8528     0.00012
bmi:1                                     -0.03944   0.02983  -1.3223     0.18605
hla1:1                                     0.00970   0.22178   0.0438     0.96510
tabac_10y1:1                               0.34881   0.20881   1.6705     0.09482
profession2:1                             -0.55385   0.27313  -2.0278     0.04258
profession3:1                             -1.39105   0.41357  -3.3635     0.00077
bdmard_lag1:1                              0.04139   0.30875   0.1341     0.89335
beta20                                     2.17963   1.05227   2.0714     0.03832
synm_tot_lag1:2                           -1.88124   0.78299  -2.4026     0.01628
Osteo_synm_tot_corr_upp_low_2_lag1:2       3.45389   0.61935   5.5766     < 2e-16
Osteo_synm_tot_corr_upp_low_2_lag2:2      -5.68746   1.05295  -5.4015     < 2e-16
sexe1:2                                   -1.71239   0.25189  -6.7982     < 2e-16
age_m0:2                                   0.01266   0.01604   0.7894     0.42989
bmi:2                                     -0.00536   0.03170  -0.1692     0.86568
hla1:2                                    -0.26445   0.24467  -1.0808     0.27977
tabac_10y1:2                               0.43831   0.23403   1.8728     0.06109
profession2:2                             -0.12582   0.32364  -0.3888     0.69745
profession3:2                             -1.11732   0.49924  -2.2380     0.02522
bdmard_lag1:2                              0.05147   0.33133   0.1553     0.87655

beta10                                    ***
synm_tot_lag1:1                           ***
Osteo_synm_tot_corr_upp_low_2_lag1:1      *
Osteo_synm_tot_corr_upp_low_2_lag2:1      ***
sexe1:1                                   ***
age_m0:1                                  ***
bmi:1
hla1:1
tabac_10y1:1                              .
profession2:1                             *
```

```
profession3:1                           ***
bdmard_lag1:1
beta20                                   *
synm_tot_lag1:2                          *
Osteo_synm_tot_corr_upp_low_2_lag1:2 ***
Osteo_synm_tot_corr_upp_low_2_lag2:2 ***
sexe1:2                                  ***
age_m0:2
bmi:2
hla1:2
tabac_10y1:2                               .
profession2:2
profession3:2                            *
bdmard_lag1:2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
     [,1] [,2] [,3] [,4]
[1,]    0    0    1    1
[2,]    0    0    1    1
[3,]    1    1    0    0
[4,]    1    1    0    0


p-value of Null model: < 0.0001
```

Although the parameter estimates differ the two models above provide the same fitted values

```
> fitted_values_gee2 <-
+   model_gee_multinomial2 |>
+   fitted.values() |>
+   _[, c(2, 3, 1)] |>
+   as.numeric()
> expect_equal(fitted_values_gee, fitted_values_gee2)
```

Hence having different reference categories is not affecting inference.

Comparison with SAS

Finally, I compared the output of the model `model_gee_independence`. According to the theory, the output from SAS and `model_gee_independence` should identical. Indeed they are.

```
> model_gee_independence |>
+     summary()
GEE FOR NOMINAL MULTINOMIAL RESPONSES
version 1.6.0 modified 2017-07-10


Link : Baseline Category Logit

Local Odds Ratios:
Structure:          independence
Homogenous scores: TRUE


call:
nomLORgee(formula = Osteo_synm_tot_corr_upp_low_2 ~ synm_tot_lag +
```

```
    Osteo_synm_tot_corr_upp_low_2_lag + sexe + age_m0 + bmi +
    hla + tabac_10y + profession + bdmard_lag, data = lesions,
    id = id_vertebra, repeated = t_new, LORstr = "independence")

Summary of residuals:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.993958 -0.036387 -0.017108  0.000000 -0.005257  0.998641

Number of Iterations: 1

Coefficients:
                                          Estimate   san.se    san.z  Pr(>|san.z|)
beta10                                    -6.52439  0.47953 -13.6057      < 2e-16
synm_tot_lag1:1                            0.05270  0.72423   0.0728      0.94199
Osteo_synm_tot_corr_upp_low_2_lag1:1       4.89730  0.20368  24.0438      < 2e-16
Osteo_synm_tot_corr_upp_low_2_lag2:1       0.43359  1.05362   0.4115      0.68069
sexe1:1                                   -0.11244  0.12254  -0.9176      0.35885
age_m0:1                                   0.06970  0.00667  10.4531      < 2e-16
bmi:1                                      0.03408  0.01391   2.4508      0.01425
hla1:1                                    -0.27415  0.11875  -2.3086      0.02097
tabac_10y1:1                               0.08950  0.11574   0.7732      0.43937
profession2:1                              0.42803  0.18911   2.2635      0.02361
profession3:1                              0.27373  0.30940   0.8847      0.37631
bdmard_lag1:1                              0.01008  0.13640   0.0739      0.94109
beta20                                    -8.70402  0.98462  -8.8400      < 2e-16
synm_tot_lag1:2                            1.93394  0.53170   3.6373      0.00028
Osteo_synm_tot_corr_upp_low_2_lag1:2       1.44341  0.63914   2.2584      0.02392
Osteo_synm_tot_corr_upp_low_2_lag2:2       6.12105  0.40832  14.9909      < 2e-16
sexe1:2                                    1.59995  0.22586   7.0840      < 2e-16
age_m0:2                                   0.05704  0.01480   3.8528      0.00012
bmi:2                                      0.03944  0.02983   1.3223      0.18605
hla1:2                                    -0.00970  0.22178  -0.0438      0.96510
tabac_10y1:2                              -0.34881  0.20881  -1.6705      0.09482
profession2:2                              0.55385  0.27313   2.0278      0.04258
profession3:2                              1.39105  0.41357   3.3635      0.00077
bdmard_lag1:2                             -0.04139  0.30875  -0.1341      0.89335

beta10                                    ***
synm_tot_lag1:1
Osteo_synm_tot_corr_upp_low_2_lag1:1 ***
Osteo_synm_tot_corr_upp_low_2_lag2:1
sexe1:1
age_m0:1                                  ***
bmi:1                                     *
hla1:1                                    *
tabac_10y1:1
profession2:1                             *
profession3:1
bdmard_lag1:1
beta20                                    ***
synm_tot_lag1:2                           ***
Osteo_synm_tot_corr_upp_low_2_lag1:2 *
Osteo_synm_tot_corr_upp_low_2_lag2:2 ***
```

```
sexe1:2                                ***
age_m0:2                               ***
bmi:2
hla1:2
tabac_10y1:2                            .
profession2:2                          *
profession3:2                          ***
bdmard_lag1:2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Local Odds Ratios Estimates:
     [,1] [,2] [,3] [,4]
[1,]    0    0    1    1
[2,]    0    0    1    1
[3,]    1    1    0    0
[4,]    1    1    0    0

p-value of Null model: < 0.0001
```