Monday, 11 March 2002

Analysis of Epileptic Seizures and Chemotheraphy.

In a clinical trial, 59 patients with epilepsy were randomized to two groups receiving either anti-epileptic drug or a placebo. The number of seizures were counted over a four-week periods. In addition a base line seizure rate was recorded for each patient.

We will first discuss the use of the **`xtgee`** command. The main components of the model have to be specified are:

- The assumed distribution of the response variable, specified in the **`family()`** option.
- The link between the response variable and its linear predictor, specified in the **`link()`** option
- The structure of the working correlation matrix, specified in the **`correlations()`** option.

Let's look at some exploratory analysis.
The following tables summarize the counts at each of the four time points

```
. summarize  counts1 counts2 counts3 counts4

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
     counts1 |      59    8.949153   14.83521          0        102
     counts2 |      59    8.355932   10.18749          0         65
     counts3 |      59    8.440678   14.14856          0         76
     counts4 |      59    7.305085   9.649467          0         63

. summarize  counts1 counts2 counts3 counts4 if  teatment==0

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
     counts1 |      28    9.357143   10.13689          0         40
     counts2 |      28    8.285714   8.164318          0         29
     counts3 |      28    8.785714   14.67262          0         76
     counts4 |      28    7.964286   7.627835          0         29

. summarize  counts1 counts2 counts3 counts4 if  teatment==1

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
     counts1 |      31    8.580645   18.24057          0        102
     counts2 |      31    8.419355   11.85966          0         65
     counts3 |      31    8.129032   13.89422          0         72
     counts4 |      31    6.709677   11.26408          0         63
```

Some graphics of the data may be useful for investigating the relation between the response and the covariates. The most useful graphical display for investigating the data is a set of graphs of individual response profiles. To include the baseline value in the graphs we duplicate first observation for each subject (week =1) and change one of the

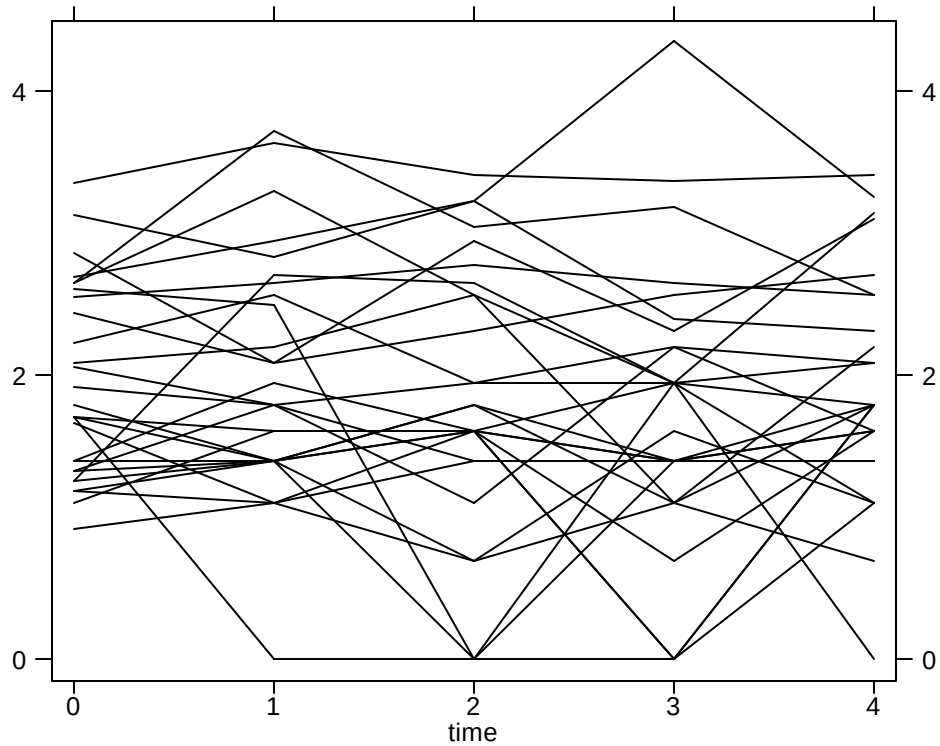
duplicates for each subject to represent the baseline measure (week = 0). Since the baseline measure represents seizure counts over an eight week period, compared with two week periods for each of the other time-points, we divide the baseline measure by 4.

```
. expand 2 if  time==1
(59 observations created)

. edit
- preserve

. sort  ID time

. qui by  ID: replace  time=0 if _n==1

. replace y = baselein/4 if  time == 0
variable y not found
r(111);

. replace count = baseline/4 if  time == 0
(57 real changes made)
```

Since we plan to fit a Poisson Model with the log link to the data, we take the log transformation before plotting the response profiles. We need to add 1 because some seizure counts are 0.

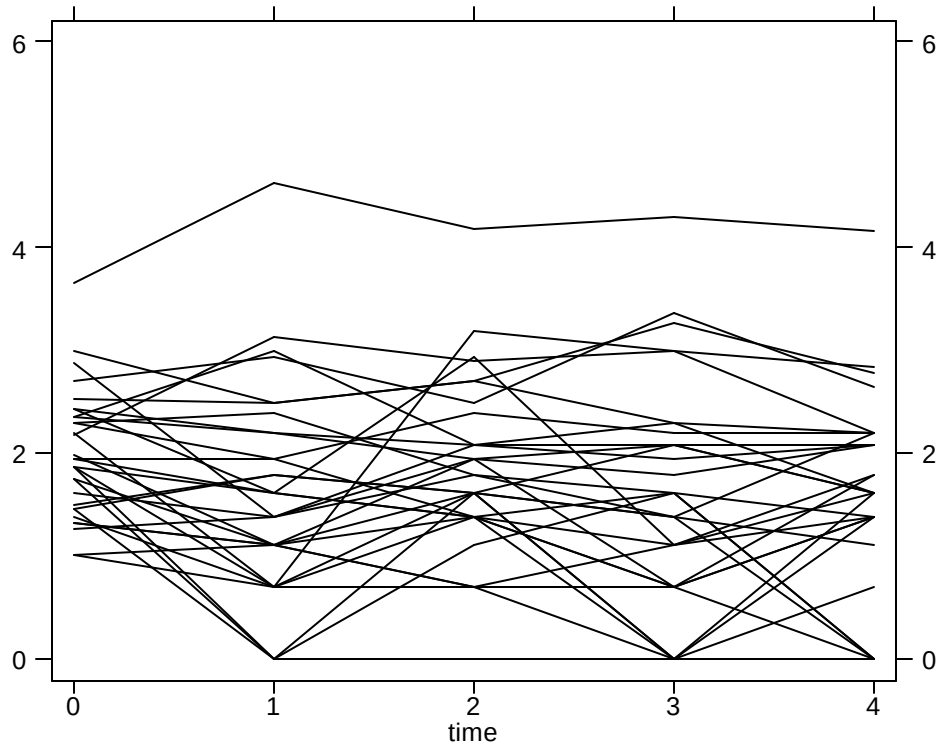
For the placebo group.

```
. graph ly  time if  teatment==0, c(L) s(i) xlab ylab rlabel
```



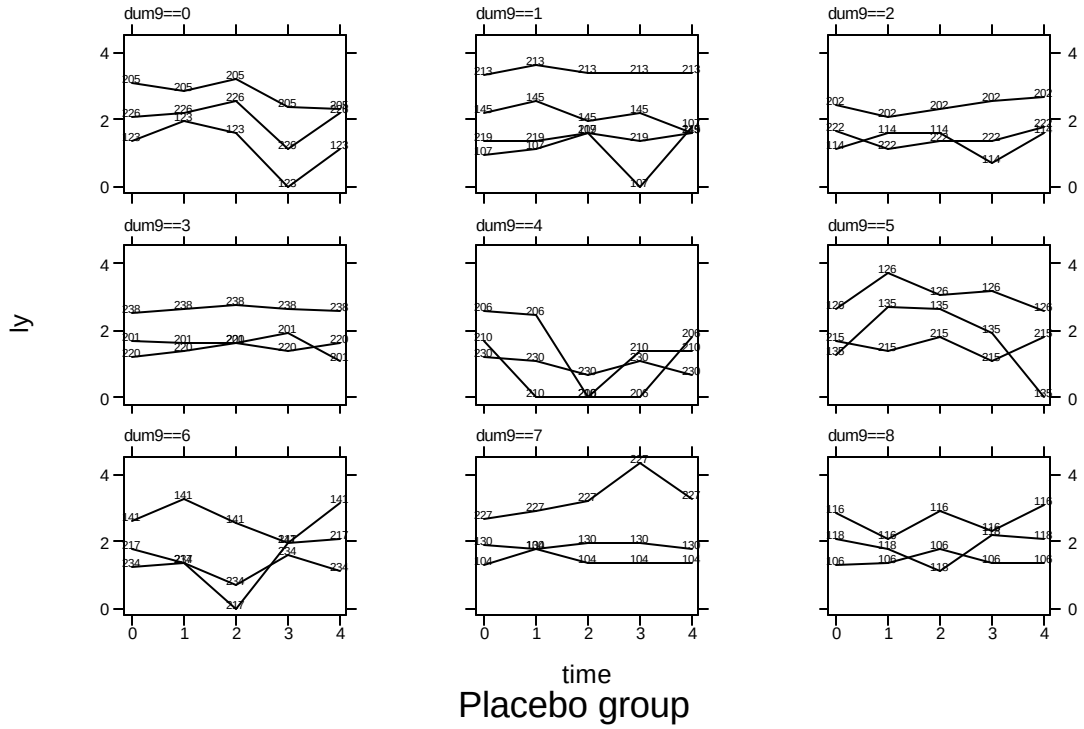
For the treatment group

```
. graph ly  time if  teatment==1, c(L) s(i) xlab ylab rlabel
```



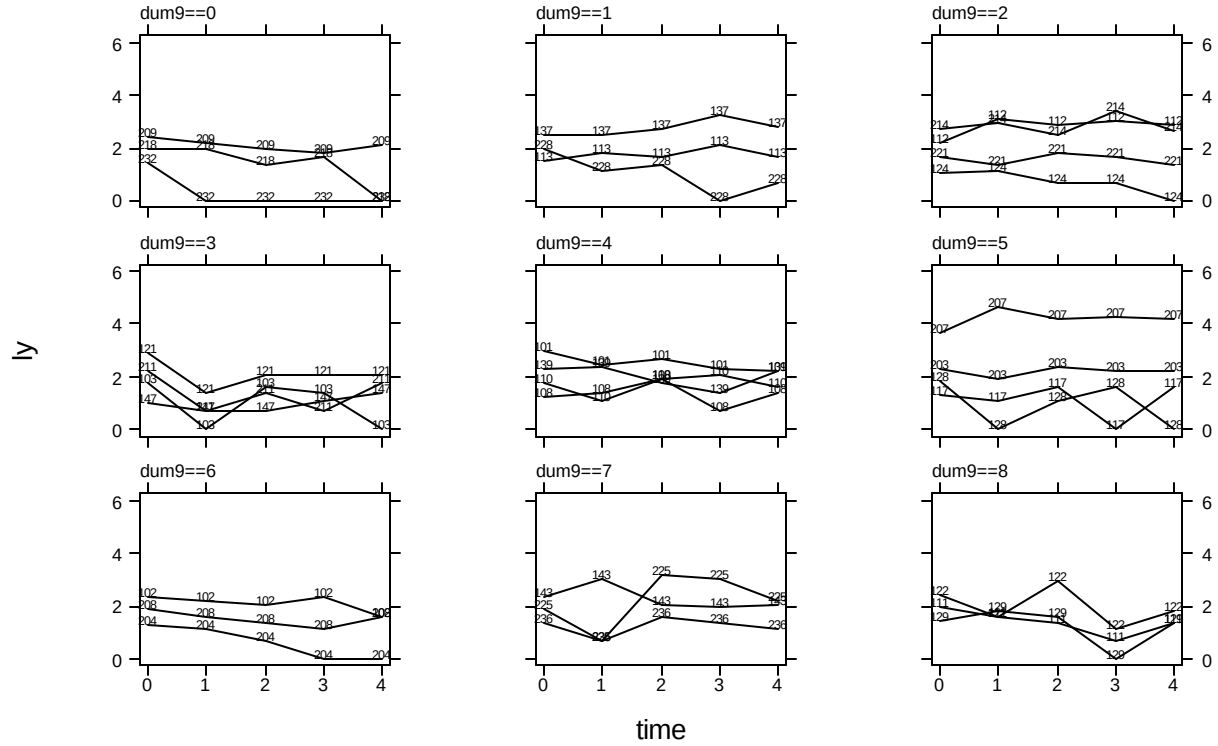
There is too much overlap between subjects; we use 9 graphs in each group. First we produce a grouping variable dum9 that splits each treatment group into nine groups. If subjects are numbered by variable **`i,`** we can apply the function **`mod(i, 9)`** to create a variable that numbers subjects from 0, …, 8, 0, …, 8, …. The graphs are clearer to read if the baseline values within a graph are not too similar. We therefore sort by baseline within each treatment group and define **i** to number subjects in this order.

```
. sort  teatment baseline ID time

. quietly by  teatment baseline ID: replace id = _n==1

. replace id = sum(id)
(294 real changes made)

. gen dum9=mod(id, 9)

. sort dum9  ID time

. graph ly  time if  teatment==0, c(L) s([ ID]) by( dum9) xlab ylab
rlabel  b1("Placebo group")
```



Placebo group

```
. graph ly  time if  teatment==1, c(L) s([ ID]) by( dum9) xlab ylab
rlabel  b1("Treatment group")
```



Treatment group

Subject 207 had more epileptic fits overall than any other subject.
It is reasonably to assume Poisson distribution for counts data. The Poisson distribution is specified in **`xtgee`** models using the option **`family(poisson)`** and the link is log.

Model with overdispersion:

```
. xtgee  counts teatment baseline Age time, i( ID) corr(exc) f(pois) l(log) sca
> le(x2)

Iteration 1: tolerance = .0183008
Iteration 2: tolerance = 2.535e-06
Iteration 3: tolerance = 1.030e-09

GEE population-averaged model                   Number of obs      =       236
Group variable:                         ID      Number of groups   =        59
Link:                                  log      Obs per group: min =         4
Family:                            Poisson                     avg =       4.0
Correlation:                  exchangeable                     max =         4
                                                Wald chi2(4)       =    194.09
Scale parameter:                  4.999686      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      counts |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    teatment |  -.1478458   .1586985    -0.93   0.352    -.4588891    .1631975
    baseline |   .0227431   .0016898    13.46   0.000     .0194312     .026055
         Age |   .0235715   .0133574     1.76   0.078    -.0026085    .0497516
        time |  -.0587233   .0350856    -1.67   0.094    -.1274898    .0100431
       _cons |   .6759401   .4581396     1.48   0.140    -.2219971    1.573877
------------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion)

. xtgee  counts teatment baseline Age time, i( ID)  corr(exc) f(pois) l(log)

Iteration 1: tolerance = .0183008
Iteration 2: tolerance = 2.535e-06
Iteration 3: tolerance = 1.030e-09

GEE population-averaged model                   Number of obs      =       236
Group variable:                         ID      Number of groups   =        59
Link:                                  log      Obs per group: min =         4
Family:                            Poisson                     avg =       4.0
Correlation:                  exchangeable                     max =         4
                                                Wald chi2(4)       =    970.41
Scale parameter:                           1    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      counts |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    teatment |  -.1478458   .0709743    -2.08   0.037     -.286953   -.0087386
    baseline |   .0227431   .0007557    30.10   0.000      .021262    .0242243
         Age |   .0235715   .0059738     3.95   0.000     .0118631      .03528
        time |  -.0587233   .0156912    -3.74   0.000    -.0894776   -.0279691
       _cons |   .6759401   .2048927     3.30   0.001     .2743578    1.077522
------------------------------------------------------------------------------
```

The estimated correlation matrix is,

```
. xtcorr

Estimated within-ID correlation matrix R:
```

```
        c1      c2      c3      c4
r1  1.0000
r2  0.4033  1.0000
r3  0.4033  0.4033  1.0000
r4  0.4033  0.4033  0.4033  1.0000
```

Treatment is significant leading to a qualitatively different conclusion about the effectiveness of the treatment. Let's use the **`robust`** option and see what happens.

```
. xtgee  counts teatment baseline Age time, i( ID)  corr(exc) f(pois) l(log)
robust

Iteration 1: tolerance = .0183008
Iteration 2: tolerance = 2.535e-06
Iteration 3: tolerance = 1.031e-09

GEE population-averaged model                   Number of obs      =       236
Group variable:                          ID     Number of groups   =        59
Link:                                   log     Obs per group: min =         4
Family:                             Poisson                     avg =       4.0
Correlation:                  exchangeable                     max =         4
                                                Wald chi2(4)       =    603.40
Scale parameter:                         1      Prob > chi2        =    0.0000

                               (standard errors adjusted for clustering on ID)
------------------------------------------------------------------------------
             |             Semi-robust
      counts |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    teatment |  -.1478458   .1701226    -0.87   0.385    -.4812799    .1855883
    baseline |   .0227431   .0012543    18.13   0.000     .0202848    .0252014
         Age |   .0235715   .0119041     1.98   0.048     .0002399    .0469031
        time |  -.0587233    .035298    -1.66   0.096    -.1279062    .0104595
       _cons |   .6759401   .3570996     1.89   0.058    -.0239622    1.375842
------------------------------------------------------------------------------
```

The estimated coefficient of treatment describes the difference in the log of the average seizure counts between the placebo and treatment groups. The negative value indicates that the treatment is more effective than the placebo in controlling the seizure rate although not significant. The exponential coefficient gives an incident rate ratio, here it represents the ratio of average seizures rates, measured as the number of seizures per two-week period, for the treated patients compared to that among the control patients. The exponential coefficient and the corresponding confidence interval can be obtained directly using the **`eform`** option in **`xtgee:`**

```
. xtgee  counts teatment baseline Age time, i( ID)  corr(exc) f(pois) l(log)
robust eform

Iteration 1: tolerance = .0183008
Iteration 2: tolerance = 2.535e-06
Iteration 3: tolerance = 1.030e-09

GEE population-averaged model                   Number of obs      =       236
Group variable:                          ID     Number of groups   =        59
Link:                                   log     Obs per group: min =         4
Family:                             Poisson                     avg =       4.0
```

```
Correlation:                  exchangeable                     max =         4
                                                   Wald chi2(4)       =    603.40
Scale parameter:                          1        Prob > chi2        =    0.0000

                               (standard errors adjusted for clustering on ID)
------------------------------------------------------------------------------
             |             Semi-robust
      counts |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    treatment |   .8625641   .1467416    -0.87   0.385     .6179919    1.203927
    baseline |   1.023004   .0012831    18.13   0.000     1.020492    1.025522
         Age |   1.023852    .012188     1.98   0.048      1.00024     1.04802
        time |   .9429676   .0332849    -1.66   0.096     .8799359    1.010514
------------------------------------------------------------------------------
```

The results suggest that there is a 14% reduction in the incidence rate of epileptic seizures in the treated group compared with the control group. According to the confidence interval, the reduction could be as 38% or there could be much as an 18% increase.

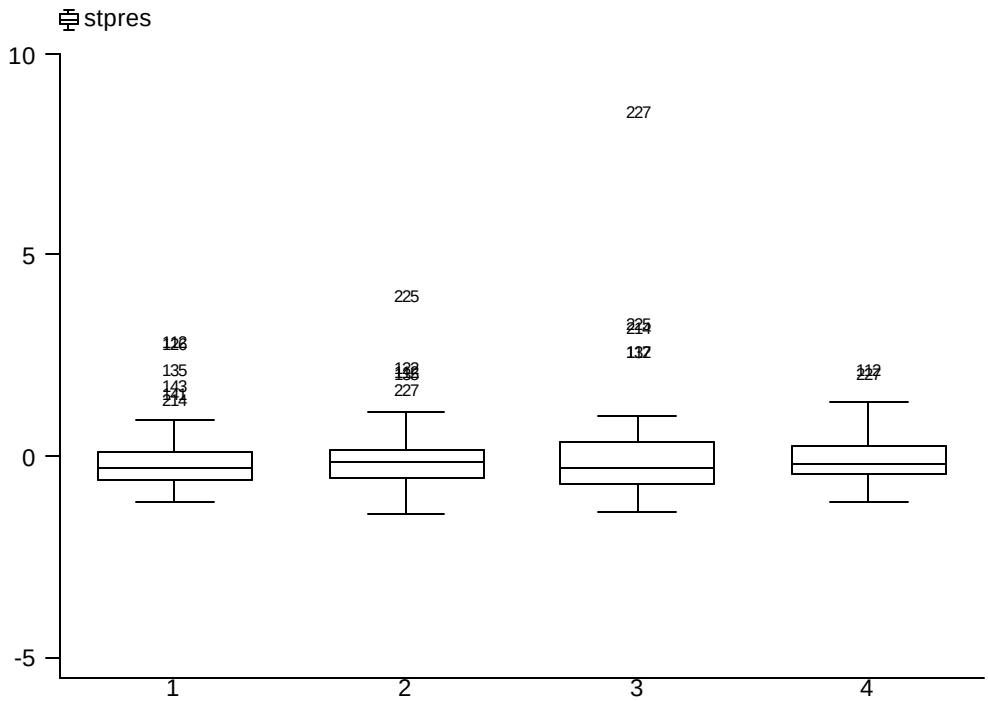
There might be interaction between baseline seizure count and treatment, a model is easily fitted using the following:

```
. gen bt =  teatment* baseline
. xtgee  counts teatment baseline Age time bt, i( ID)  corr(exc) f(pois) l(log)
eform

Iteration 1: tolerance = .01738938
Iteration 2: tolerance = .00001068
Iteration 3: tolerance = 1.147e-08

GEE population-averaged model                   Number of obs      =       236
Group variable:                         ID      Number of groups   =        59
Link:                                  log      Obs per group: min =         4
Family:                            Poisson                     avg =       4.0
Correlation:                  exchangeable                     max =         4
                                                   Wald chi2(5)       =    983.15
Scale parameter:                          1        Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      counts |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    treatment |   .7735594   .0879598    -2.26   0.024     .6190209    .9666784
    baseline |   1.021598   .0013912    15.69   0.000     1.018875    1.024329
         Age |   1.025602   .0062896     4.12   0.000     1.013348    1.038004
        time |   .9429676     .01479    -3.74   0.000     .9144207    .9724057
          bt |      1.002   .0016398     1.22   0.222     .9987908    1.005219
```

There is no evidence of an interaction. One can try other correlation structures.
We now look at the standardized Pearson's residuals of the previous model separately for each week. This can be done by using the **`predict`** command to obtain estimated counts and computing the standardized Pearson's residuals.

```
. predict pred, xb
. replace pred = exp(pred)
(236 real changes made)
. gen pres = (counts - pred)/sqrt(pred)
. gen stpres = pres/sqrt(e(chi2_dis))
. sort time
. graph stpres, box s([ID]) by(time) ylab
```

The resulting graph is,



Here the subject 227 is clearly an influential observation in week 3.

You can also a random effects model. Random effects is fitted using restricted maximum likelihood estimation, using **`xtpois`** (you can explore more about this in STATA help).