A Handbook of Statistical Analyses Using R

Brian S. Everitt and Torsten Hothorn
CHAPTER 11

Analysing Longitudinal Data II – Generalised Estimation Equations: Treating Respiratory Illness and Epileptic Seizures

11.1 Introduction

11.2 Generalised Estimating Equations

11.3 Analysis Using R

11.3.1 Beat the Blues Revisited

To use the gee function, package gee (Carey et al., 2006) has to be installed and attached:

```R
R> library("gee")
```

The gee function is used in a similar way to the lme function met in Chapter 10, with the addition of the features of the glm function that specify the appropriate error distribution for the response and the implied link function, and an argument to specify the structure of the working correlation matrix. Here we will fit an independence structure and then an exchangeable structure. The R code for fitting generalised estimation equations to the BtheB_long data (as constructed in Chapter 10, with identity working correlation matrix is as follows (note that the gee function assumes the rows of the data.frame BtheB_long to be ordered with respect to subjects)

```R
R> osub <- order(as.integer(BtheB_long$subject))
R> BtheB_long <- BtheB_long[osub,]
R> btb_gee <- gee(bdi ~ bdi.pre + treatment + length + drug,
+       data = BtheB_long, id = subject, family = gaussian,
+       corstr = "independence")
```

and with exchangeable correlation matrix

```R
R> btb_gee1 <- gee(bdi ~ bdi.pre + treatment + length + drug,
+       data = BtheB_long, id = subject, family = gaussian,
+       corstr = "exchangeable")
```

The summary method can be used to inspect the fitted models; the results are shown in Figures 11.1 and 11.2
R> summary(btb_gee)

GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
Link: Identity
Variance to Mean Relation: Gaussian
Correlation Structure: Independent

Call:
gee(formula = bdi ~ bdi.pre + treatment + length + drug, id = subject,
data = BtheB_long, family = gaussian, corstr = "independence")

Summary of Residuals:

Min 1Q Median 3Q Max
-21.6497810 -5.8485100 0.1131663 5.5838383 28.1871039

Coefficients:

Estimate  Naive S.E.  Naive z  Robust S.E.  Robust z
(Intercept)  3.5686314  1.4833349  2.405816  2.26947617  1.5724472
bdi.pre  0.5818494  0.0563904  10.318235  0.09156455  6.3545274
treatmentBtheB -3.2372285  1.1295569  -2.865928  1.77459534  -1.8242066
length>6m  1.4577182  1.1380277  1.280916  1.48255866  0.9832449
drugYes  -3.7412982  1.1763904  -3.179667  1.78271179  -2.0986557

Estimated Scale Parameter:  79.25813
Number of Iterations:  1

Working Correlation
[1,] 1 0 0 0
[2,] 0 1 0 0
[3,] 0 0 1 0
[4,] 0 0 0 1

Figure 11.1  R output of the summary method for the btb_gee model.

11.3.2 Respiratory Illness

The baseline status, i.e., the status for month == 0, will enter the models as an explanatory variable and thus we have to rearrange the data.frame respiratory in order to create a new variable baseline:

R> data("respiratory", package = "HSAUR")
R> resp <- subset(respiratory, month > "0")
R> resp$baseline <- rep(subset(respiratory, month == "0")$status, rep(4, 111))
R> resp$nstat <- as.numeric(resp$status == "good")

The new variable nstat is simply a dummy coding for a poor respiratory status. Now we can use the data resp to fit a logistic regression model and GEE models with an independent and an exchangeable correlation structure as follows;
**ANALYSIS USING R**

```
R> summary(btb_gee1)

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gg S-function, version 4.13 modified 98/01/27 (1998)

Model:                           
Link: Identity                  
Variance to Mean Relation: Gaussian
Correlation Structure: Exchangeable

Call:                            
gee(formula = bdi ~ bdi.pre + treatment + length + drug, id = subject, 
    data = BtheB_long, family = gaussian, corstr = "exchangeable")

Summary of Residuals:           
     Min  1Q Median  3Q Max  
-23.955980 -6.643864 -1.109741 4.257688 25.452310

Coefficients:                    
                          Estimate Naive S.E.  Naive z Robust S.E. 
(Intercept) 3.0231602       2.30390185 1.31219140   2.23204410
bdi.pre 0.6479276          0.08228567 7.87412417  0.08351405
treatmentBtheB -2.1692863  1.76642861 -1.22806339  1.73614385
length>6m -0.1112910       1.73091679 -0.06429596  1.55092705
drugYes -2.9995608         1.82569913 -1.64296559  1.73155411

Robust z                      
(Intercept) 1.3544357
bdi.pre 7.5830666

treatmentBtheB -1.2494854
length>6m -0.7175777

drugYes -1.732940

Estimated Scale Parameter:  81.7349
Number of Iterations:  5

Working Correlation

[1,] 1.00000000 0.6757951 0.6757951 0.6757951 
[2,] 0.6757951 1.0000000 0.6757951 0.6757951 
[3,] 0.6757951 0.6757951 1.0000000 0.6757951 
[4,] 0.6757951 0.6757951 0.6757951  1.0000000

Figure 11.2  R output of the summary method for the btb_gee1 model.
```

```
R> resp_glm <- glm(status ~ centre + treatment + sex + baseline +
    + age, data = resp, family = "binomial")
R> resp_gee1 <- gee(nstat ~ centre + treatment + sex + baseline +
    + age, data = resp, family = "binomial", id = subject,
    + corstr = "independence", scale.fix = TRUE, scale.value = 1)
R> resp_gee2 <- gee(nstat ~ centre + treatment + sex + baseline +
    + age, data = resp, family = "binomial", id = subject,
    + corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)
```
R> summary(resp_glm)

Call:
glm(formula = status ~ centre + treatment + sex + baseline + age, family = "binomial", data = resp)

Deviance Residuals:
    Min      1Q  Median      3Q     Max
-2.3146  -0.8551   0.4336   0.8953   1.9246

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.900171   0.337653  -2.666  0.00768 **
   centre2     0.671601   0.239567   2.803  0.00506 **
treatment1    1.299216   0.236841   5.486  4.12e-08 ***
   sexmale     0.119244   0.294671   0.405  0.68572
   baseline1   1.882029   0.241290   7.800  6.20e-15 ***
   age        -0.018166   0.008864  -2.049  0.04043 *

---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 608.93  on 443  degrees of freedom
Residual deviance: 483.22  on 438  degrees of freedom
AIC: 495.22

Number of Fisher Scoring iterations: 4

Figure 11.3  R output of the `summary` method for the `resp_glm` model.
ANALYSIS USING R

R> summary(resp_gee1)

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
Link: Logit
Variance to Mean Relation: Binomial
Correlation Structure: Independent

Call:
gee(formula = nstat ~ centre + treatment + sex + baseline + age,
    id = subject, data = resp, family = "binomial", corstr = "independence",
    scale.fix = TRUE, scale.value = 1)

Summary of Residuals:
    Min  1Q Median  3Q Max
-0.93134415 -0.30623174  0.08973552  0.33018952  0.84307712

Coefficients:
                   Estimate  Naive S.E.  Naive z
(Intercept) -0.90017133   0.337653052 -2.665965
centre2 0.67160098   0.239566599  2.803400
treatmenttreatment 1.29921589   0.236841017  5.485603
sexmale 0.11924365   0.294671045  0.404667
baselinegood 1.88202860   0.241290221  7.799854
age -0.01816588   0.008864403 -2.049306

                   Robust S.E.  Robust z
(Intercept)  0.46032700 -1.9555041
centre2 0.35681913  1.8218899
treatmenttreatment 0.35077797  3.7038127
sexmale 0.44320235  0.2690501
baselinegood 0.35005152  5.3764332
age 0.01300426 -1.3969169

Estimated Scale Parameter: 1
Number of Iterations: 1

Working Correlation

[1,] 1 0 0 0
[2,] 0 1 0 0
[3,] 0 0 1 0
[4,] 0 0 0 1

---

Figure 11.4  R output of the summary method for the resp_gee1 model.
**R**> summary(resp_gee2)

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gg S-function, version 4.13 modified 98/01/27 (1998)

Model:
Link: Logit
Variance to Mean Relation: Binomial
Correlation Structure: Exchangeable

Call:
`gee(formula = nstat ~ centre + treatment + sex + baseline + age,`
`   id = subject, data = resp, family = "binomial", corstr = "exchangeable",`
`   scale.fix = TRUE, scale.value = 1)`

Summary of Residuals:
Min 1Q Median 3Q Max
-0.93134415 -0.30623174 0.08973552 0.33018952 0.84307712

Coefficients:

|                | Estimate | Naive S.E. | Naive z  
|----------------|----------|------------|----------
| (Intercept)    | -0.9002  | 0.4785     | -1.8814  |
| centre2        | 0.6716   | 0.3395     | 1.9783   |
| treatment     | 1.2992   | 0.3356     | 3.8712   |
| sexmale        | 0.1192   | 0.4176     | 0.2856   |
| baselinegood  | 1.8820   | 0.3419     | 5.5044   |
| age            | -0.0182  | 0.0126     | -1.4462  |

Robust S.E. | Robust z
|-------------|-------------
| (Intercept) | 0.4603      | -1.9555    |
| centre2     | 0.3568      | 1.8821     |
| treatment   | 0.3508      | 3.7038     |
| sexmale     | 0.4432      | 0.2690     |
| baseline    | 0.3505      | 5.3764     |
| age         | 0.0130      | -1.3969    |

Estimated Scale Parameter: 1
Number of Iterations: 1

Working Correlation

```
[1,] 1.0000000 0.3359883 0.3359883 0.3359883  
[2,] 0.3359883 1.0000000 0.3359883 0.3359883  
[3,] 0.3359883 0.3359883 1.0000000 0.3359883  
[4,] 0.3359883 0.3359883 0.3359883 1.0000000  
```

**Figure 11.5** R output of the `summary` method for the `resp_gee2` model.
The estimated treatment effect taken from the exchangeable structure GEE model is 1.299 which, using the robust standard errors, has an associated 95% confidence interval.

\[
R > \text{se} \leftarrow \text{summary(resp\_gee2)$coefficients["treatmenttreatment", + "Robust S.E."\]}
R > \text{coef(resp\_gee2)["treatmenttreatment"] + c(-1, 1) * se * qnorm(0.975)}
\]

\[\begin{bmatrix} 0.6117037 \\ 1.9867281 \end{bmatrix}\]

These values reflect effects on the log-odds scale. Interpretation becomes simpler if we exponentiate the values to get the effects in terms of odds. This gives a treatment effect of 3.666 and a 95% confidence interval of

\[
R > \exp(\text{coef(resp\_gee2)["treatmenttreatment"] + + c(-1, 1) * se * qnorm(0.975)})
\]

\[\begin{bmatrix} 1.843570 \\ 7.291637 \end{bmatrix}\]

The odds of achieving a ‘good’ respiratory status with the active treatment is between about twice and seven times the corresponding odds for the placebo.

### 11.3.3 Epilepsy

Moving on to the count data in epilepsy from Table ??, we begin by calculating the means and variances of the number of seizures for all treatment / period interactions.

\[
R > \text{data("epilepsy", package = "HSAUR")}
R > \text{itp <- interaction(epilepsy$treatment, epilepsy$period)}
R > \text{tapply(epilepsy$seizure.rate, itp, mean)}
\]

\[
\begin{array}{cccc}
\text{placebo.1} & \text{Progabide.1} & \text{placebo.2} & \text{Progabide.2} \\
9.357143 & 8.580645 & 8.285714 & 8.419355 \\
\text{placebo.3} & \text{Progabide.3} & \text{placebo.4} & \text{Progabide.4} \\
8.129032 & 7.964286 & 6.709677 \\
\end{array}
\]

\[
R > \text{tapply(epilepsy$seizure.rate, itp, var)}
\]

\[
\begin{array}{cccc}
\text{placebo.1} & \text{Progabide.1} & \text{placebo.2} & \text{Progabide.2} \\
102.75661 & 332.71828 & 66.65608 & 140.65161 \\
\text{placebo.3} & \text{Progabide.3} & \text{placebo.4} & \text{Progabide.4} \\
193.04946 & 58.18386 & 126.87957 \\
\end{array}
\]

Some of the variances are considerably larger than the corresponding means, which for a Poisson variable may suggest that overdispersion may be a problem, see Chapter ??.

We can now fit a Poisson regression model to the data assuming independence using the \texttt{glm} function. We also use the GEE approach to fit an independence structure, followed by an exchangeable structure using the following R code:

\[
R > \text{data("epilepsy", package = "HSAUR")}
R > \text{itp <- interaction(epilepsy$treatment, epilepsy$period)}
R > \text{tapply(epilepsy$seizure.rate, itp, mean)}
R > \text{tapply(epilepsy$seizure.rate, itp, var)}
\]
R> layout(matrix(1:2, nrow = 1))
R> ylim <- range(epilepsy$seizure.rate)
R> placebo <- subset(epilepsy, treatment == "placebo")
R> progabide <- subset(epilepsy, treatment == "Progabide")
R> boxplot(seizure.rate ~ period, data = placebo,
+   ylab = "Number of seizures",
+   xlab = "Period", ylim = ylim, main = "Placebo")
R> boxplot(seizure.rate ~ period, data = progabide,
+   main = "Progabide", ylab = "Number of seizures",
+   xlab = "Period", ylim = ylim)

![Boxplots of numbers of seizures in each two-week period post randomisation for placebo and active treatments.](image)

**Figure 11.6** Boxplots of numbers of seizures in each two-week period post randomisation for placebo and active treatments.

R> per <- rep(log(2), nrow(epilepsy))
R> epilepsy$period <- as.numeric(epilepsy$period)
R> fm <- seizure.rate ~ base + age + treatment + offset(per)
R> epilepsy_glm <- glm(fm, data = epilepsy, family = "poisson")
R> epilepsy_gee1 <- gee(fm, data = epilepsy, family = "poisson",
+   id = subject, corstr = "independence", scale.fix = TRUE,
+   scale.value = 1)
R> epilepsy_gee2 <- gee(fm, data = epilepsy, family = "poisson",
+   id = subject, corstr = "exchangeable", scale.fix = TRUE,
+   scale.value = 1)
R> epilepsy_gee3 <- gee(fm, data = epilepsy, family = "poisson",
+   id = subject, corstr = "independence", scale.fix = TRUE,
+   scale.value = 1)
ANALYSIS USING R

```r
R> layout(matrix(1:2, nrow = 1))
R> ylim <- range(log(epilepsy$seizure.rate + 1))
R> boxplot(log(seizure.rate + 1) ~ period, data = placebo,
+   main = "Placebo", ylab = "Log number of seizures",
+   xlab = "Period", ylim = ylim)
R> boxplot(log(seizure.rate + 1) ~ period, data = progabide,
+   main = "Progabide", ylab = "Log number of seizures",
+   xlab = "Period", ylim = ylim)
```

![Boxplots of log numbers of seizures in each two-week period post randomisation for placebo and active treatments.](image)

**Figure 11.7** Boxplots of log of numbers of seizures in each two-week period post randomisation for placebo and active treatments.

As usual we inspect the fitted models using the `summary` method, the results are given in Figures 11.8, 11.9, 11.10, and 11.11.
```r
R> summary(epilepsy_glm)

Call:
glm(formula = fm, family = "poisson", data = epilepsy)

Deviance Residuals:
  Min        1Q    Median        3Q       Max
-4.4360    -1.4034    -0.5029     0.4842   12.3223

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.1306156  0.1356191  -0.963   0.3355
base         0.0226517  0.0005093  44.476  < 2e-16 ***
age          0.0227401  0.0040240   5.651   1.59e-08 ***
treatmentProgabide -0.1527009  0.0478051  -3.194   0.0014 **

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2521.75  on 235  degrees of freedom
Residual deviance: 958.46  on 232  degrees of freedom
AIC: 1732.5

Number of Fisher Scoring iterations: 5
```

**Figure 11.8** R output of the `summary` method for the `epilepsy_glm` model.
ANALYSIS USING R

R> summary(epilepsy_gee1)

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
  Link: Logarithm
  Variance to Mean Relation: Poisson
  Correlation Structure: Independent

Call:
  gee(formula = fm, id = subject, data = epilepsy, family = "poisson",
       corstr = "independence", scale.fix = TRUE, scale.value = 1)

Summary of Residuals:

         .Min   1Q  Median     3Q  Max
-4.9195387 0.1808059 1.7073405 4.8850644 69.9658560

Coefficients:

                     Estimate  Naive S.E.  Naive z
(Intercept)        -0.13061561 0.1356191185 -0.9631062
base               0.02265174 0.0005093011 44.4761250
age                0.02274013 0.0040239970  5.6511312
base 0.02265174 0.0005093011 44.4761250
age 0.02274013 0.0040239970  5.6511312
age 0.02274013 0.0040239970  5.6511312
treatmentProgabide -0.15270095 0.0478051054 -3.1942393
Robust S.E. Robust z
(Intercept) 0.365148155 -0.3577058
base 0.001235664 18.3316325
age 0.011580405 1.9636736
treatmentProgabide 0.171108915 -0.8924196

Estimated Scale Parameter: 1
Number of Iterations: 1

Working Correlation

[1,] 1 0 0 0
[2,] 0 1 0 0
[3,] 0 0 1 0
[4,] 0 0 0 1

Figure 11.9  R output of the summary method for the epilepsy_gee1 model.
R> summary(epilepsy_gee2)

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gg S-function, version 4.13 modified 98/01/27 (1998)

Model:
Link: Logarithm
Variance to Mean Relation: Poisson
Correlation Structure: Exchangeable

Call:
gee(formula = fm, id = subject, data = epilepsy, family = "poisson",
corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)

Summary of Residuals:
                           Min     1Q   Median     3Q    Max
-4.9195387  0.1808059  1.7073405  4.8850644 69.9658560

Coefficients:             Estimate  Naive S.E.  Naive z
(Intercept)             -0.13061561  0.2004416507 -0.6516391
base                     0.02265174  0.0007527342 30.0926122
age                      0.02274013  0.0059473665  3.8235638
treatmentProgabide     -0.15270095  0.0706547450 -2.1612270

Robust S.E.  Robust z
(Intercept) 0.365148155 -0.3577058
base 0.001235664 18.3316325
age 0.011580405 1.9636736
treatmentProgabide 0.171108915 -0.8924196

Estimated Scale Parameter: 1
Number of Iterations: 1

Working Correlation
[1,] 1.0000000 0.3948033 0.3948033 0.3948033
[2,] 0.3948033 1.0000000 0.3948033 0.3948033
[3,] 0.3948033 0.3948033 1.0000000 0.3948033
[4,] 0.3948033 0.3948033 0.3948033 1.0000000

Figure 11.10  R output of the summary method for the epilepsy_gee2 model.
ANALYSIS USING R

R> summary(epilepsy_gee3)

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gg S-function, version 4.13 modified 98/01/27 (1998)

Model:
Link: Logarithm
Variance to Mean Relation: Poisson
Correlation Structure: Exchangeable

Call:
gee(formula = fm, id = subject, data = epilepsy, family = "poisson",
corstr = "exchangeable", scale.fix = FALSE, scale.value = 1)

Summary of Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>-4.9195</td>
<td>0.1808</td>
<td>1.7073</td>
<td>4.8850</td>
<td>69.97</td>
</tr>
</tbody>
</table>

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Naive S.E.</th>
<th>Naive z</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.1306</td>
<td>0.4522</td>
<td>-0.2888</td>
</tr>
<tr>
<td>base</td>
<td>0.0226</td>
<td>0.0017</td>
<td>13.33</td>
</tr>
<tr>
<td>age</td>
<td>0.0227</td>
<td>0.0135</td>
<td>1.69</td>
</tr>
<tr>
<td>treatment</td>
<td>-0.1527</td>
<td>0.1594</td>
<td>-0.96</td>
</tr>
</tbody>
</table>

Robust S.E. Robust z

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Robust S.E.</th>
<th>Robust z</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.3651</td>
<td>0.0012</td>
<td>18.33</td>
</tr>
<tr>
<td>base</td>
<td>0.0012</td>
<td>0.0015</td>
<td>13.31</td>
</tr>
<tr>
<td>age</td>
<td>0.0011</td>
<td>1.96</td>
<td>9.64</td>
</tr>
<tr>
<td>treatment</td>
<td>0.1711</td>
<td>0.89</td>
<td>19.19</td>
</tr>
</tbody>
</table>

Estimated Scale Parameter: 5.08961
Number of Iterations: 1

Working Correlation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>1.0000000</td>
<td>0.3948033</td>
<td>0.3948033</td>
<td>0.3948033</td>
</tr>
<tr>
<td>[2,]</td>
<td>0.3948033</td>
<td>1.0000000</td>
<td>0.3948033</td>
<td>0.3948033</td>
</tr>
<tr>
<td>[3,]</td>
<td>0.3948033</td>
<td>0.3948033</td>
<td>1.0000000</td>
<td>0.3948033</td>
</tr>
<tr>
<td>[4,]</td>
<td>0.3948033</td>
<td>0.3948033</td>
<td>0.3948033</td>
<td>1.0000000</td>
</tr>
</tbody>
</table>

Figure 11.11  R output of the summary method for the epilepsy_gee3 model.
Bibliography