



# **A Handbook of Statistical Analyses Using R**

---

Brian S. Everitt and Torsten Hothorn



## 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> 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> 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> 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

### 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;

```
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(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.
(Intercept)	3.5686314	1.4833349	2.405816	2.26947617
bdi.pre	0.5818494	0.0563904	10.318235	0.09156455
treatmentBtheB	-3.2372285	1.1295569	-2.865928	1.77459534
length>6m	1.4577182	1.1380277	1.280916	1.48255866
drugYes	-3.7412982	1.1766321	-3.179667	1.78271179

	Robust z
(Intercept)	1.5724472
bdi.pre	6.3545274
treatmentBtheB	-1.8242066
length>6m	0.9832449
drugYes	-2.0986557

```
Estimated Scale Parameter: 79.25813
```

```
Number of Iterations: 1
```

```
Working Correlation
```

	[,1]	[,2]	[,3]	[,4]
[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.

---

```
R> summary(btb_gee1)
```

*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: 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	7.7583066
treatmentBtheB	-1.2494854
length>6m	-0.0717577
drugYes	-1.7322940

Estimated Scale Parameter: 81.7349  
 Number of Iterations: 5

Working Correlation

	[,1]	[,2]	[,3]	[,4]
[1,]	1.0000000	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> 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 **
treatmenttreatment	1.299216	0.236841	5.486	4.12e-08 ***
sexmale	0.119244	0.294671	0.405	0.68572
baselinegood	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.

---

```
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	1.8821889
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]	[,2]	[,3]	[,4]
[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
gee 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)	-0.90017133	0.47846344	-1.8813796
centre2	0.67160098	0.33947230	1.9783676
treatmenttreatment	1.29921589	0.33561008	3.8712064
sexmale	0.11924365	0.41755678	0.2855747
baselinegood	1.88202860	0.34191472	5.5043802
age	-0.01816588	0.01256110	-1.4462014

	Robust S.E.	Robust z
(Intercept)	0.46032700	-1.9555041
centre2	0.35681913	1.8821889
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]	[,2]	[,3]	[,4]
[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> se <- summary(resp_gee2)$coefficients["treatmenttreatment",
+                                         "Robust S.E."]
R> coef(resp_gee2)["treatmenttreatment"] +
+     c(-1, 1) * se * qnorm(0.975)
[1] 0.6117037 1.9867281
```

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(coef(resp_gee2)["treatmenttreatment"] +
+     c(-1, 1) * se * qnorm(0.975))
[1] 1.843570 7.291637
```

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> data("epilepsy", package = "HSAUR")
R> itp <- interaction(epilepsy$treatment, epilepsy$period)
R> tapply(epilepsy$seizure.rate, itp, mean)

  placebo.1 Progabide.1 placebo.2 Progabide.2 placebo.3
  9.357143   8.580645   8.285714   8.419355   8.785714
Progabide.3 placebo.4 Progabide.4
  8.129032   7.964286   6.709677

R> tapply(epilepsy$seizure.rate, itp, var)

  placebo.1 Progabide.1 placebo.2 Progabide.2 placebo.3
  102.75661  332.71828   66.65608  140.65161  215.28571
Progabide.3 placebo.4 Progabide.4
  193.04946   58.18386  126.87957
```

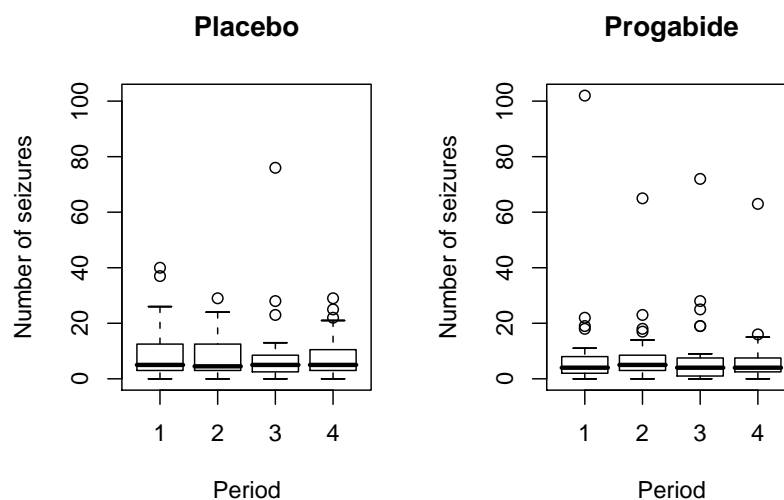
Some of the variances are considerably larger than the corresponding means, which for a Poisson variable may suggest that overdispersion may be a problem, see Chapter ?. We can now fit a Poisson regression model to the data assuming independence using the `glm` function. We also use the GEE approach to fit an independence structure, followed by an exchangeable structure using the following R code:

```
R> per <- rep(log(2), nrow(epilepsy))
R> epilepsy$period <- as.numeric(epilepsy$period)
```

```

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)

```



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

```

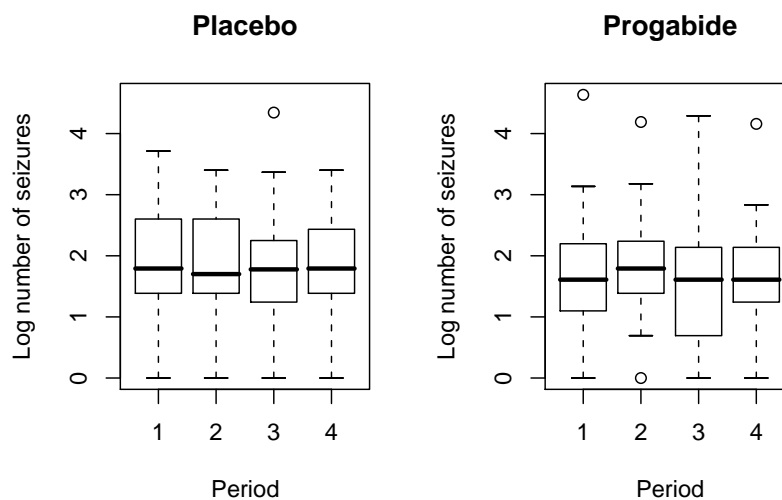
R> fm <- seizure.rate ~ base + age + treatment + offset(per)
R> epilepsy_glm <- glm(fm, data = epilepsy, family = "poisson")
R> epilepsy_gee1 <- gee(fm, data = epilepsy, family = "poisson",
+   id = subject, corstr = "independence", scale.fix = TRUE,
+   scale.value = 1)
R> epilepsy_gee2 <- gee(fm, data = 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 = "exchangeable", scale.fix = FALSE,
+   scale.value = 1)

```

```

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)

```



**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> 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	0.33549
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	0.00140 **

```
---
```

```
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.

---

```
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
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]	[,2]	[,3]	[,4]
[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*  
*gee 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	-0.651639
base	0.02265174	0.0007527342	30.092612
age	0.02274013	0.0059473665	3.823564
treatmentProgabide	-0.15270095	0.0706547450	-2.161227

	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]	[,2]	[,3]	[,4]
[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.

---

```
R> summary(epilepsy_gee3)

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:      Exchangeable

Call:
gee(formula = fm, id = subject, data = epilepsy, family = "poisson",
    corstr = "exchangeable", scale.fix = FALSE, 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 0.452199543 -0.2888451
base          0.02265174 0.001698180 13.3388301
age           0.02274013 0.013417353  1.6948302
treatmentProgabide -0.15270095 0.159398225 -0.9579840
              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:  5.089608
Number of Iterations:  1

Working Correlation
      [,1]      [,2]      [,3]      [,4]
[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 11.11** R output of the `summary` method for the `epilepsy_gee3` model.



---

## Bibliography

---

Carey, V. J., Lumley, T., and Ripley, B. D. (2006), *gee: Generalized Estimation Equation Solver*, URL <http://CRAN.R-project.org>, R package version 4.13-11.