10. Marginal models for non-normal responses (GEE)

Sonja Greven

Summer Term 2015

Overview Chapter 10 - Marginal models for non-normal responses (GEE)

The marginal model

- In generalized linear mixed models it is assumed that the association between $Y_{i1}, \ldots, Y_{in_i}$ is explained by random effects. Thus, the modeling of the association is linked to the modeling of the expected value $(g(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i)$.

- In marginal models the expected value and the association are modeled separately:

$\rightarrow$ The expected value is modeled by $g(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta}$.

The marginal model: general principle

1. The marginal expectation $\mu_{ij} = \mathsf{E}(Y_{ij})$ is linked to the covariates via a known link function g:

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}.$$

2. The marginal variance $\mathsf{Var}(Y_{ij})$ depends on the marginal mean through the known variance function v:

$$\mathsf{Var}(Y_{ij}) = \phi v(\mu_{ij}).$$

3. The correlation between $Y_{i1}, \ldots, Y_{in_i}$ is a a known function ρ of the marginal means and an additional parameter vector $\boldsymbol{\alpha}$:

$$\mathsf{Corr}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \boldsymbol{\alpha}).$$

The marginal model:
example with continuous response

- Expected value $\mu_{ij} = \mathsf{E}(Y_{ij})$:

$$\mu_{ij} = \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}.$$

- Variance:

$$\mathsf{Var}(Y_{ij}) = \phi v(\mu_{ij}) = \phi,$$

  i.e. variance remains the same over time (possibly unrealistic).

- Association:

$$\mathsf{Corr}(Y_{ij}, Y_{ik}) = \alpha^{|k-j|}$$

  with $0 \leq \alpha \leq 1$.

The marginal model: example with count variable

- Expected value $\mu_{ij} = \mathsf{E}(Y_{ij})$:

$$\log(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}$$

- Variance:

$$\mathsf{Var}(Y_{ij}) = \phi \mu_{ij}$$

- Association:

$$\mathsf{Corr}(Y_{ij}, Y_{ik}) = \alpha_{jk}, \quad \text{unstructured}$$

The marginal model: example with binary response

- Expected value $\mu_{ij} = \mathsf{E}(Y_{ij})$:

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta}.$$

- Variance ($\phi = 1$):

$$\mathsf{Var}(Y_{ij}) = \mu_{ij}(1-\mu_{ij}).$$

- Association:

$$\log \mathsf{OR}(Y_{ij}, Y_{ik}) = \alpha_{jk}, \quad \text{unstructured},$$

where

$$\mathsf{OR}(Y_{ij}, Y_{ik}) = \frac{P(Y_{ij}=1, Y_{ik}=1)P(Y_{ij}=0, Y_{ik}=0)}{P(Y_{ij}=1, Y_{ik}=0)P(Y_{ij}=0, Y_{ik}=1)}.$$

Side note: why use the ORs instead of the correlations?

- $$\begin{aligned} \text{Corr}(Y_{ij}, Y_{ik}) &= \frac{\text{E}(Y_{ij}Y_{ik}) - \text{E}(Y_{ij})\text{E}(Y_{ik})}{\sqrt{\text{Var}(Y_{ij})\text{Var}(Y_{ik})}} \\ &= (P(Y_{ij} = 1, Y_{ik} = 1) - \mu_{ij}\mu_{ik}) / \sqrt{\mu_{ij}(1 - \mu_{ij})\mu_{ik}(1 - \mu_{ik})} \end{aligned}$$

- $-1 \leq \text{Corr}(Y_{ij}, Y_{ik}) \leq 1$ is fulfilled if this complicated constraint holds:

$$\max(0, \mu_{ij} + \mu_{ik} - 1) \leq P(Y_{ij} = 1, Y_{ik} = 1) \leq \min(\mu_{ij}, \mu_{ik}).$$

  In particular, we cannot reasonably assume that the correlation is independent of the covariates.

- The ORs $\in (0, \infty)$ are not constrained by the means.

The marginal model

For normally distributed responses

$$\mathbf{Y}_i \sim \mathcal{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i)$$

the model is fully specified by the first two moments (multivariate normal distribution). This results in a (relatively) simple expression for the likelihood.

This is different in the case of non-normal responses!

For example: binary response ($Y_{ij} = 0$ or $Y_{ij} = 1$).

Example binary response

- When assuming
$$\log \mathsf{OR}(Y_{ij}, Y_{ik}) = \alpha_{jk}$$
for $j \neq k$ and $i = 1, \ldots, N$, where α_{jk} are parameters, the joint distribution of $Y_{i1}, \ldots, Y_{in_i}$ is not fully specified.

- To specify the joint distribution of $Y_{i1}, \ldots, Y_{in_i}$, one needs $2^{n_i} - n_i - 1$ association parameters (for all 2- to n_i-size subsets of $Y_{i1}, \ldots, Y_{in_i}$)!

The marginal model

- For this reason, ML-based approaches are very difficult to apply (e.g. Bahadur model). Typically, simplifying assumptions are necessary.

- Most of these approaches are computationally intensive! They are rarely used in practice.

- **Viable alternative:** Generalized estimating equations (GEEs).
  In GEEs, the focus lies on modeling the mean.

Overview Chapter 10 - Marginal models for non-normal responses (GEE)

Generalized estimating equations

- Approach of Liang and Zeger (1986)
- The focus lies on modeling the mean (first moment).
- Applicable to both normally (e.g. if the covariance structure is unknown) and non-normally distributed data.
- Implementation in R: R-packages `gee` and `geepack`.
- Implementation in SAS: `PROC GENMOD`.

Generalized estimating equations

Reminder: GLS estimator

The weighted least squares (GLS) estimator

$$\widehat{\boldsymbol{\beta}}_{\mathbf{V}_i^{-1}} = \left\{ \sum_{i=1}^{N} (\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i) \right\}^{-1} \sum_{i=1}^{N} (\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i)$$

minimizes the criterion

$$\sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}).$$

Generalized estimating equations

It can be shown that the GLS estimator fulfills the score equations

$$\sum_{i=1}^{N} \mathbf{X}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0},$$

where $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}) = \mathbf{X}_i \boldsymbol{\beta}$.

Idea of the GEE estimator: minimize

$$\sum_{i=1}^{N} (\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})),$$

with working covariance $\mathbf{V}_i$ and

$$\mu_{ij} = \mu_{ij}(\boldsymbol{\beta}) = g^{-1}(\mathbf{x}_{ij}^T \boldsymbol{\beta}).$$

Generalized estimating equations

If there is a minimum of the criterion

$$\sum_{i=1}^{N} (\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})),$$

it can be shown that it fulfills the score equations

$$\sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

with

$$\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \partial \mu_{i1} / \partial \beta_1 & \dots & \partial \mu_{i1} / \partial \beta_p \\ \vdots & \vdots & \vdots \\ \partial \mu_{in_i} / \partial \beta_1 & \dots & \partial \mu_{in_i} / \partial \beta_p \end{pmatrix}.$$

Summary: score equations in GLS/GLM/GEE

- GLS:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{X}_i^T \mathbf{V}_i^{-1} (\boldsymbol{y}_i - \mathbf{X}_i \boldsymbol{\beta}) = \mathbf{0}.$$

- GLM:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{\partial \theta_i}{\partial \boldsymbol{\beta}} \left[y_i - \psi'(\theta_i) \right] = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} v_i^{-1} (y_i - \mu_i) = \mathbf{0}.$$

- GEE:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T \mathbf{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}.$$

Overview Chapter 10 - Marginal models for non-normal responses (GEE)

11.3. Estimation

The equations

$$\sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T \mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) \quad = \quad \mathbf{0}$$

are called **generalized estimating equations** and can be written as

$$\sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T (\mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2})^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

with

$$\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2},$$

where $\mathbf{A}_i = \text{diag}(\phi v(\mu_{i1}), \ldots, \phi v(\mu_{in_i}))$ and $\mathbf{R}_i$ is the correlation matrix.

Generalized estimating equations

$$\sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T (\mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2})^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}.$$

- The $p \times n_i$ matrix $\left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)^T$ depends only on $\boldsymbol{\beta}$.
- $\mathbf{A}_i = \mathbf{A}_i(\boldsymbol{\beta})$ results from the marginal model.
- But $\boldsymbol{\beta}$ contains no information about $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\alpha})$.
  $\longrightarrow$ One has to make assumptions.

Generalized estimating equations: estimation algorithm

1. Calculate initial estimates $\widehat{\boldsymbol{\beta}}^{(1)}$ for $\boldsymbol{\beta}$, e.g. with a GLM under the independence assumption.

2. Update $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\phi}$ from the current $\widehat{\boldsymbol{\beta}}^{(t)}$, further explanations next slide.

3. Derive $\mathbf{A}_i(\widehat{\boldsymbol{\beta}})$, $\mathbf{R}_i(\widehat{\boldsymbol{\alpha}})$, $\widehat{\mathbf{V}}_i$ and $\widehat{\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}}$.

4. Update $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}}^{(t+1)} = \widehat{\boldsymbol{\beta}}^{(t)} - \left[\sum_{i=1}^{N} \left(\widehat{\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}} \right)^T \widehat{\mathbf{V}}_i^{-1} \left(\widehat{\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}} \right) \right]^{-1} \times \left[\sum_{i=1}^{N} \left(\widehat{\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}} \right)^T \widehat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i) \right]$$

5. Repeat 2-4 until convergence.

Estimation of α and ϕ

For given $\widehat{\boldsymbol{\beta}}$, the Pearson residuals

$$\tilde{r}_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{v(\hat{\mu}_{ij})}}$$

can be calculated. Estimators for $\boldsymbol{\alpha}$ are obtained using moment based estimators for various common assumptions on the structure:

Structure	Corr(Y_{ij}, Y_{ik})	Estimator
Independence	0	
Exchangeable	α	$\widehat{\alpha} = \frac{1}{N\hat{\phi}} \sum_{i=1}^{N} \frac{1}{n_i(n_i-1)} \sum_{j \neq k} \tilde{r}_{ij}\tilde{r}_{ik}$
AR(1)	$\alpha^{\lvert j-k \rvert}$	$\widehat{\alpha} = \frac{1}{N\hat{\phi}} \sum_{i=1}^{N} \frac{1}{n_i-1} \sum_{j \leq n_i-1} \tilde{r}_{ij}\tilde{r}_{i,j+1}$
Unstructured	α_{jk}	$\widehat{\alpha}_{jk} = \frac{1}{N\hat{\phi}} \sum_{i=1}^{N} \tilde{r}_{ij}\tilde{r}_{ik}$

Similarly, ϕ is estimated (if $\phi \neq 1$):

$$\widehat{\phi} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{r}_{ij}^2.$$

Overview Chapter 10 - Marginal models for non-normal responses (GEE)

Properties and inference

1. $\widehat{\boldsymbol{\beta}}$ is a **consistent estimator** for $\boldsymbol{\beta}$ as $N \to \infty$. This will also apply if the structural assumption for $\mathbf{R}_i(\boldsymbol{\alpha})$ is incorrect: $\widehat{\boldsymbol{\beta}}$ is a robust estimator. Only the marginal expectation must be specified correctly.

2. $\widehat{\boldsymbol{\beta}}$ has an asymptotically multivariate normal distribution with covariance matrix
$$\text{Cov}(\widehat{\boldsymbol{\beta}}) = \mathbf{B}^{-1}\mathbf{M}\mathbf{B}^{-1},$$
where
$$\begin{aligned}
\mathbf{B} &= \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T \mathbf{V}_i^{-1} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right) \\
\mathbf{M} &= \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T \mathbf{V}_i^{-1} \text{Cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right).
\end{aligned}$$

Properties and inference

- **Note:** In which way will $\mathbf{B}$ and $\mathbf{M}$ simplify if the identity function is selected as link function?

- For estimation of $\mathbf{B}$ and $\mathbf{M}$, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ und ϕ are replaced by $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\phi}$.

- Cov$(\mathbf{Y}_i)$ is estimated by

$$(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)^T.$$

  $\longrightarrow$ Sandwich estimator of Cov$(\widehat{\boldsymbol{\beta}})$.

- Consistent for $N \rightarrow \infty$.

Model-based estimator for Cov($\widehat{\boldsymbol{\beta}}$)

The model-based estimator for Cov($\widehat{\boldsymbol{\beta}}$) is

$$\text{Cov}(\widehat{\boldsymbol{\beta}}) = \mathbf{B}^{-1}.$$

It is preferable to model the covariance and use the model-based estimator instead of the robust estimator when:

- the sample size N is small (relative to n_i),
- there are few subjects for each combination of covariates,
- the data are strongly unbalanced.

GEE: Conclusion

Advantages:

- Often nearly as precise and efficient as ML estimators if Var$(\boldsymbol{Y}_i)$ is reasonably approximated by the working covariance $\boldsymbol{V}_i$.
- For multivariate normal responses the GEE estimator is very similar to the GLS estimator. Therefore, GLS can be seen as a special case of GEE.
- GEE, in contrast to ML approaches, is robust against misspecification of the covariance structure. $\widehat{\boldsymbol{\beta}}$ and the sandwich estimator of Cov$(\widehat{\boldsymbol{\beta}})$ are both consistent for $N \to \infty$ even if Var$(\boldsymbol{Y}_i)$ is incorrectly specified.

But...

- The efficiency is higher when the association is correctly specified.
- The robustness is an asymptotic property.

Overview Chapter 10 - Marginal models for non-normal responses (GEE)

Example: Toenail data

- Y_{ij}: binary outcome (severe infection yes/no)
- t_{ij}: time
- T_{ij}: treatment (0,1)
- Consider this model for Y_{ij} with three possible correlation structures:

$$
\begin{aligned}
\text{logit}(\mathsf{E}(Y_{ij})) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} \\
\text{Var}(Y_{ij}) &= \phi \pi_{ij}(1-\pi_{ij}) \\
\text{Corr}(Y_{ij}, Y_{ik}) &\in \{0, \alpha, \alpha_{jk}\}.
\end{aligned}
$$

- Note: Is β_1 necessary?

Example: Toenail data

```
> gee(Response ~ Month * Treatment, id = ID, corstr="independence", family=binomial)
Coefficients:
                   Estimate Naive S.E.      Naive z Robust S.E.     Robust z
(Intercept)     -0.5566272625 0.11139546 -4.996857562  0.17117080 -3.251882106
Month           -0.1703077870 0.02414727 -7.052879577  0.02916250 -5.839958270
Treatment       -0.0005816626 0.15963276 -0.003643754  0.25084786 -0.002318786
Month:Treatment -0.0672216208 0.03836181 -1.752305936  0.05211553 -1.289857677
> gee(Response ~ Month * Treatment, id = ID, corstr="exchangeable", family=binomial)
Coefficients:
                  Estimate Naive S.E.     Naive z Robust S.E.    Robust z
(Intercept)     -0.581850825 0.14027487 -4.14793340  0.17204948 -3.38188069
Month           -0.171274123 0.02103731 -8.14144496  0.02999742 -5.70962885
Treatment        0.007190544 0.19493800  0.03688631  0.25945870  0.02771364
Month:Treatment -0.077723656 0.03571174 -2.17641757  0.05410892 -1.43642956
> gee(Response ~ Month * Treatment, id = ID, corstr="unstructured", family=binomial)
Coefficients:
                 Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept)     -0.69928288 0.17026346 -4.1070637  0.16700042 -4.1873122
Month           -0.14135905 0.02652237 -5.3298049  0.02700176 -5.2351789
Treatment        0.03760836 0.24106235  0.1560109  0.24385339  0.1542253
Month:Treatment -0.08283103 0.04279448 -1.9355538  0.04798388 -1.7262261
```

Example: Toenail data

- Robust and model-based standard errors are closest for the unstructured working correlation. Thus, the unstructured correlation may be closest to the structure of the data and resulting estimates most efficient.

- The working unstructured correlation is estimated to decrease with time distance.

- ϕ is estimated to be close to 1.

- An alternative would specify the correlation structure using odds ratios (cf. p. 7). This leads to **alternating logistic regression** discussed e.g. in Molenberghs & Verbeke, 2005, Ch. 8, and implemented in SAS `proc genmod`. The R package `orth` is unfortunately no longer available.

Example: Toenail data

GEE β estimates are smaller in absolute value than those from the GLMM in Chapter 9, as expected. The GLMM assumes conditional independence, which may not completely capture the correlation structure in the data.

```
> gee(Response ~ Month * Treatment, id = ID, corstr="unstructured", family=binomial)
Coefficients:
                   Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept)     -0.69928288 0.17026346 -4.1070637  0.16700042 -4.1873122
Month           -0.14135905 0.02652237 -5.3298049  0.02700176 -5.2351789
Treatment        0.03760836 0.24106235  0.1560109  0.24385339  0.1542253
Month:Treatment -0.08283103 0.04279448 -1.9355538  0.04798388 -1.7262261

> glmer(Response ~ Month * Treatment + (1 | ID), family = binomial, nAGQ=25)
Fixed effects:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.61464    0.43280  -3.731 0.000191 ***
Month           -0.39083    0.04435  -8.812  < 2e-16 ***
Treatment       -0.16003    0.58275  -0.275 0.783620
Month:Treatment -0.13675    0.06798  -2.012 0.044245 *
```

Example: Epileptic seizures

- Y_{ij}: count outcome (number of seizures)
- T_{ij} = length of observation period: $T_{i1} = 8$, $T_{i2} = T_{i3} = T_{i4} = T_{i5} = 2$.
- $B_i = 0$ for placebo, $B_i = 1$ for progabide.
- $F_{ij} = 0$ for baseline, $F_{ij} = 1$ else.
- Consider the model:

$$\begin{aligned} \log(\mathsf{E}(Y_{ij})) = \log(\mu_{ij}) &= \log T_{ij} + \beta_1 + \beta_2 F_{ij} + \beta_3 B_i + \beta_4 B_i F_{ij} \\ \mathsf{Var}(Y_{ij}) &= \phi \mu_{ij} \\ \mathsf{Corr}(Y_{ij}, Y_{ik}) &= \alpha, \quad j \neq k. \end{aligned}$$

Example: Epileptic seizures

```
> library(gee)
> gee(count ~ offset(log(weeks)) + followup * group,
           id = id,  corstr="exchangeable", family=poisson)
Coefficients:
                      Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept)         1.34760922  0.1511851  8.9136359   0.1573571  8.5640166
followupTRUE        0.11079814  0.1547038  0.7161956   0.1160997  0.9543358
group               0.02651461  0.2072721  0.1279217   0.2218539  0.1195138
followupTRUE:group -0.10368067  0.2199500 -0.4713830   0.2136100 -0.4853736

Estimated Scale Parameter:  19.70269
```

Diggle et al (2002), p. 164, discuss that patient 49 is very unusual, with an extremely high seizure count of 151 in 8 weeks at baseline and a doubling to 302 seizures in 8 weeks after treatment. Without this patient, there is a modest indication of a treatment benefit ($\widehat{\beta}_4 = -0.30\ (0.17)$).
$\widehat{\phi}$ drops from 19.4 to 10.4 and $\widehat{\alpha}$ from 0.78 to 0.60.

Example: Epileptic seizures

Note that the estimates are almost the same as those from the GLMM with only a random intercept.

```
> glmer(count ~ offset(log(weeks)) + followup * group
                 + (1 | id), family = poisson, nAGQ = 20)
Fixed effects:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.03265    0.15254   6.769 1.29e-11 ***
followupTRUE       0.11080    0.04672   2.371   0.0177 *
group             -0.02385    0.21047  -0.113   0.9098
followupTRUE:group -0.10368    0.06482  -1.599   0.1097
```

Poisson GLMM and GEE

This example illustrates an important special case:

- In a log-linear mixed model where $\boldsymbol{z}_{ij}$ contains a subset of the variables in $\boldsymbol{x}_{ij}$, the $\boldsymbol{\beta}$ parameters for the variables in $\boldsymbol{x}_{ij}$ that are **not** in $\boldsymbol{z}_{ij}$ have the same interpretation as in a marginal model.

- In particular, $\boldsymbol{\beta}$ parameters apart from the intercept have the same interpretation in the marginal model as in a GLMM with only a random intercept (Diggle et al, 2002, p. 137).

Note also that we found in the GLMM that we should include an additional random effect for the change from baseline ($b_{i2}F_{ij}$).

Overview Chapter 10 - Marginal models for non-normal responses (GEE)

Comparison GLMM vs. GEE

- In the marginal model 1) the expected value and 2) the association between the measurements of a subject are modeled separately.
- In marginal models the covariate effects can be interpreted on the population level.
- In the GLMM the measurements of a subject are assumed to be independent conditional on the random effects.
- In the GLMM inference is on the individual level, fixed effects correspond to effects conditional on the subject.