Ziegler, Kastner, Grömping, Blettner:

The Generalized Estimating Equations in the Past Ten Years: An Overview and A Biomedical Application

Sonderforschungsbereich 386, Paper 24 (1996)

The Generalized Estimating Equations in the Past Ten Years: An Overview and A Biomedical Application

A. Ziegler * C. Kastner † U. Grömping ‡ M. Blettner §

April 29, 1996

Abstract

The Generalized Estimating Equations (GEE) proposed by Liang and Zeger (1986) have found considerable attention in the last years and several extensions have been proposed. This paper will give a more intuitive description how GEE have been developed during the last years. Additionally we will describe the advantages and disadvantages of the different parametrisations that have been proposed in the literature. We will also give a brief review of the literature available on this topic.

Keywords: Generalized Estimating Equations, Marginal Models, Correlated Data Analysis

1 Introduction

The primary goal of the analysis is in most regression models to investigate the influence of certain covariates on the response variable. Theoretical results for estimating parameters in regression models are available mainly for continuous and normal distributed response variables. However, in praxis the response variable is often binary or categorical and only recently some theoretical work has been published on regression models for this situation.

Generalized linear models (GLM) as described for example by Nelder and Wedderburn (1972) and McCullagh and Nelder (1989) are regression models to analyse continuous or discrete response variables. The association between the response variable and the covariables is given by the so-called link function. GLM assume that the observations are independent and do not consider any correlation between the outcome of the n observations. Marginal models,

*Institute of Medical Biometry, Philipps–University of Marburg, Bunsenstr. 3, 35033 Marburg, Germany, ziegler@mailer.uni-marburg.de

†Institute of Statistics, LMU München, Ludwistr. 33, 80539 München, Germany, kchris@stat.uni-muenchen.de

‡Department of Statistics, University of Dortmund, Vogelpothsweg 77, 42277 Dortmund, Germany, ulrike@omega.statistik.uni-dortmund.de

§German Cancer Research Center Heidelberg, Department of Epidemiology, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany, maria.blettner@dkfz-heidelberg.de

conditional models and random effects models are extensions of the GLM for correlated data.

In the marginal model, the primary interest of the analysis is to model the marginal expectations of the response variable given the covariables. Here, the correlation - or more general the association - between the outcome variables is modelled separately and is regarded as nuisance parameter. The major goal is to investigate the effect of the covariables in the population on the response variable. Including the correlation structure in estimating the effects mainly yields different variance estimation. Marginal models have been introduced first by Zeger, Liang and Self (1985), Liang and Zeger (1986) and Zeger and Liang (1986). In these articles the authors give a detailed definition of the approach and describe several applications. Their approach is termed Generalized Estimating Equations (GEE) and can be interpreted as a synthesis of the Feasible Generalized Least Squares (FGLS) approach (Greene, 1993) and the GLM.

Several procedures have been developed to estimate parameters in marginal models. Here, we will describe first the approach from Liang and Zeger, and then give an example and explain how to use this approach for analysing a data set with dependent observations. We will also give some recommendations for the application of this methods and a brief literature review.

2 The Generalized Estimating Equations of Order 1 (GEE1)

Let y_{it} be a vector of responses from n clusters, e.g. families or periods, with T observations for the ith cluster, $i = 1, \dots, n$. For each y_{it} several covariates $\boldsymbol{x}_{it}$ are available, where the first element of $\boldsymbol{x}_{it}$ is 1 to allow the inclusion of an intercept. The data can be summarised to the vector $\boldsymbol{y}_i$ and the matrix $\boldsymbol{X}_i = (\boldsymbol{x}'_{i1}, \dots, \boldsymbol{x}'_{iT})'$. The method can be extended to unequal cluster sizes T_i. The pairs $(\boldsymbol{y}_i, \boldsymbol{X}_i)$ are assumed to be independent identically distributed (iid).

We will first describe models for $\mathbb{E}(y_{it}|\boldsymbol{x}_{it})$. It is necessary to find a method that can deal with the association between the T observations of the cluster i. For linear models, this method is FGLS. Suppose $\mathbb{E}(y_{it}|\boldsymbol{X}_i) = \mathbb{E}(y_{it}|\boldsymbol{x}_{it}) = \boldsymbol{x}'_{it}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the unknown $p \times 1$ parameter vector of interest. Furthermore, assume that the conditional variance matrix of $\boldsymbol{y}_i$ given on $\boldsymbol{X}_i$ is known and given by $\mathbb{C}ov(\boldsymbol{y}_i|\boldsymbol{X}_i) = \boldsymbol{V}_i$. Then, the general multivariate linear regression model estimator can be obtained via the estimating equations (score equations)

$$\boldsymbol{u}(\boldsymbol{\beta}) = \frac{1}{n}\boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{0}, \tag{1}$$

where $\boldsymbol{X}$ and $\boldsymbol{y}$ are the stacked $\boldsymbol{X}_i$ matrices and $\boldsymbol{y}_i$ vectors, respectively. $\boldsymbol{V}$ is the block diagonal matrix of the $\boldsymbol{V}_i$, $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}) = \boldsymbol{X}'_i\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ is defined analogously to $\boldsymbol{y}$.

The estimator is unbiased and under the assumption of normality finite normal distributed, otherwise asymptotically normal distributed. Its covariance matrix is given by the inverse Fisher information matrix. For this estimator the Gauß–Markov–Theorem holds.

If $\mathbb{C}ov(\boldsymbol{y}_i|\boldsymbol{X}_i) = \boldsymbol{\Omega}_i \neq \boldsymbol{V}_i$, the estimator still remains unbiased. However, instead of the Fisher information matrix the robust variance matrix with the socalled sandwich form of White (1982), Gourieroux and Monfort (1984) or Liang and Zeger (1986) has to be used. This variance matrix has the form

$$\mathbb{V}(\hat{\beta}) = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{\Omega}\boldsymbol{V}^{-1}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}, \qquad (2)$$

where $\boldsymbol{\Omega}$ is defined analogously to $\boldsymbol{V}$. An estimator of (2) can be obtained by replacing $\boldsymbol{\Omega}_i$ by $\hat{\boldsymbol{\Omega}}_i = (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'$, where $\hat{\boldsymbol{\mu}}_i = \boldsymbol{X}_i\hat{\boldsymbol{\beta}}$. Note that $\hat{\boldsymbol{\Omega}}_i$ is not an estimator for $\boldsymbol{\Omega}_i$.

Instead of using a fixed covariance matrix $\boldsymbol{V}_i$, a model for $\boldsymbol{V}_i = \boldsymbol{V}_i(\boldsymbol{\alpha})$ depending on an association parameter $\boldsymbol{\alpha}$ can be used. This approach is called FGLS (Greene, 1993). The estimator $\hat{\boldsymbol{\beta}}$ is obtained in a two step procedure: In the first step the association parameter $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ and in a second step $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ is estimated.

This approach can also be used to model dependencies within clusters in linear models. However, the linear model is inadequate in many applications. For independent observations, the GLM allows flexibility in modelling mean and variance structures. In GLM, the mean structure is given by $\mathbb{E}(y_{it}|\boldsymbol{x}_{it}) = \mu_{it} = g(\boldsymbol{x}_{it}'\boldsymbol{\beta})$, where g is a non-linear response function. g^{-1} is termed link function. An important property of the GLM is the functional relation between mean and variance: $v_{it} = \mathbb{V}(y_{it}|\boldsymbol{x}_{it}) = h(\mu_{it})$. h is called variance function. In general, an assumption of the distribution motivates the link and the variance function of the GLM.

For independent observations, the parameter vector $\boldsymbol{\beta}$ is estimated using the maximum likelihood method: The distribution—say the Binomial or Poisson distribution—determines the likelihood equations (score equations) that are given by derivatives of the log-likelihood function with respect to $\boldsymbol{\beta}$. The score equations have the form

$$\boldsymbol{u}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{D}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \frac{1}{n}\boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{0}\,, \qquad (3)$$

where $\boldsymbol{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}'$ is the diagonal matrix of the first derivatives and $\boldsymbol{V}_i$ is the diagonal matrix of the variances $\boldsymbol{V}_i = \mathrm{diag}(v_{it})$. (3) are called independence estimation equations (IEE).

A solution of (3) exists only for the linear model with normal distributed response variables—in all other situations, they have to be solved iteratively. The estimator $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal distributed with covariance matrix $\mathbb{C}ov(\hat{\boldsymbol{\beta}}) = (\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}$.

For correlated observations the true variance matrix cannot have a diagonal form. For correlated data, Zeger et al. (1985) proposed the use of the robust variance estimation of White (1982):

$$\widehat{\mathbb{V}(\hat{\boldsymbol{\beta}})} = \left(\sum_{i=1}^{n}\hat{\boldsymbol{D}}_i'\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{D}}_i\right)^{-1}\left(\sum_{i=1}^{n}\hat{\boldsymbol{D}}_i'\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{\Omega}}_i\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{D}}_i\right)\left(\sum_{i=1}^{n}\hat{\boldsymbol{D}}_i'\hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{D}}_i\right)^{-1} \qquad (4)$$

In this equation, $\hat{\boldsymbol{\Omega}}$ is the block diagonal matrix of $\hat{\boldsymbol{\Omega}}_i = (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'$. $\hat{\boldsymbol{\mu}}_i$ is defined by the link function of the GLM.

The estimation will not be very efficient, because of the diagonal form of $\boldsymbol{V}_i$. An approach that has been proposed by Liang and Zeger (1986) and Zeger and Liang (1986) allows a more efficient estimation by combining the GLM and the

FGLS procedures: First, consider a model for the expectation and the variance using the GLM approach. In this case $\boldsymbol{V}_i$ is not necessarily a diagonal matrix but a covariance matrix which is closer to the true covariance matrix $\boldsymbol{\Omega}_i$ then the diagonal matrix. The association (correlation) is not of interest here. For an estimator $\hat{\boldsymbol{R}}_i$ of the correlation matrix of $\boldsymbol{y}_i$ conditional on $\boldsymbol{X}_i$, the estimator for $\boldsymbol{V}_i$ is given in the form

$$\hat{\boldsymbol{V}}_i = \hat{\boldsymbol{A}}_i^{1/2} \hat{\boldsymbol{R}}_i \hat{\boldsymbol{A}}_i^{1/2}, \tag{5}$$

with $\hat{\boldsymbol{A}}_i^{-1/2}$, the estimated inverse square root of the diagonal matrix of the variances v_{it}. $\hat{\boldsymbol{R}}_i$ should be a positive definite $T \times T$ matrix that describes well the association-structure. For estimating this 'working correlation matrix', Liang and Zeger (1986) used the method of moments. The choice of the working correlation matrix $\boldsymbol{R}_i$, has been discussed for example by Liang and Zeger (1986) in some detail. If the identy matrix is used for the working correlation matrix, (5) is reduced to a diagonal matrix for the variances, and the estimation equations are given by the IEE (3).

The interpretation of the estimation is difficult, if the working correlation matrix is not well specified—that means if it is not close to the true correlation matrix—as a function of $\boldsymbol{\alpha}$ (Crowder, 1995).

With the estimated working correlation matrix $\hat{\boldsymbol{R}}_i$ and the diagonal matrices $\hat{\boldsymbol{A}}_i$, the Generalized Estimating Equations (GEE1) have the form

$$\boldsymbol{u}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{D}_i' \hat{\boldsymbol{V}}_i^{-1} \Big(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \Big) = \boldsymbol{0}. \tag{6}$$

The term "generalized" is somehow misleading, however it is justified considering that Liang and Zeger (1986) have developed the equation system (6) from the GLM respectively IEE.

GEE1 means that only first order moments, i.e. mean structure, are estimated consistent. Note that (6) is similar to the FGLS estimator where in the first step the variance matrix $\hat{\boldsymbol{V}}_i$ and in the second step the parameter vector $\boldsymbol{\beta}$ are estimated.

If $\boldsymbol{\beta}$ is estimated using equation (6), $\hat{\boldsymbol{\beta}}$ is under suitable regularity conditions consistent, if $\mu_{it} = \mathbb{E}(y_{it}|\boldsymbol{x}_{it}) = \mathbb{E}(y_{it}|\boldsymbol{X}_i)$ is specified correctly. The GEE1 estimator is asymptotically normal distributed. The variance can be estimated consistently with the robust estimator (4) and $\hat{\boldsymbol{V}}_i$ as in (5); see e.g. Liang and Zeger (1986), Zeger and Liang (1986), Gourieroux and Monfort (1984), Rotnitzky (1988) and Ziegler (1994b). If $\boldsymbol{V}_i$ is specified correctly, i.e. $\boldsymbol{\Omega}_i = \boldsymbol{V}_i$, it follows that $\hat{\boldsymbol{\beta}}$ is efficient in the sense of Rao–Cramér
(see Gourieroux and Monfort, 1984; Ziegler, 1994b).

Most estimators proposed by Liang and Zeger (1986) for the correlation structure $\boldsymbol{R}_i$ can be developed using estimation equations (see Crowder, 1995; Ziegler, 1994b). It follows that additionally to the estimation equations for $\boldsymbol{\beta}$ a second equation system can be introduced for $\boldsymbol{\alpha}$. The general form of this estimation equation system is (Prentice, 1988)

$$\boldsymbol{u}(\boldsymbol{\alpha}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{E}_i' \boldsymbol{W}_i^{-1} \Big(\boldsymbol{z}_i - \boldsymbol{\varrho}_i(\boldsymbol{\alpha}) \Big) = \boldsymbol{0}. \tag{7}$$

In (6) the expectation μ_i of $\boldsymbol{y}_i$, is given as a function of the parameters $\boldsymbol{\beta}$. In (7) the vector form $\boldsymbol{\varrho}_i(\boldsymbol{\alpha})$ of the correlation matrix $\boldsymbol{R}_i(\boldsymbol{\alpha})$ is given as a function of the parameters of association $\boldsymbol{\alpha}$. $\boldsymbol{z}_i$ is defined analogous to the response vector $\boldsymbol{y}_i$. Note that the product of the Pearson–residuals $z_{itt'}$ do not only include observations but also parameters. $\boldsymbol{E}_i$ is the matrix including the first derivatives of $\boldsymbol{\varrho}_i(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$. $\boldsymbol{W}_i^{-1}$ can be interpreted as the inverse of the covariance matrix of $\boldsymbol{z}_i$.

The advantage of using (7) is that non-linear correlation structures can be estimated. Like the link function in the GLM, we can define the association with the explanatory variables $\boldsymbol{X}_i$ in the form $\boldsymbol{\varrho}_i(\boldsymbol{\alpha}) = \boldsymbol{\varrho}_i(\boldsymbol{X}_i, \boldsymbol{\alpha})$ (Prentice, 1988). However, it is not straight forward to define a reasonable function to model the association between the correlation structure $\boldsymbol{\varrho}_i(\boldsymbol{X}_i, \boldsymbol{\alpha})$ and the covariables $\boldsymbol{X}_i$ (see Lipsitz, Laird and Harrington, 1991; Lipsitz, Fitzmaurice, Orav and Laird, 1994; Ziegler, Kastner, Grömping and Blettner, 1996; Ziegler, 1995). The problem is that the covariables can be continuous but the correlations are restricted to the interval [-1;1]. Therefore it is necessary to define restrictions for the correlation structure which should be non-linear functions analogous to the well-known link function. An example for such a function is given by the inverse of Fisher's z–transformation (Lipsitz et al., 1991). The transformation has a similar interpretation as the link function in GLM and we will call it "link function for the association".

3 Generalized Estimating Equations for Estimation of Mean and Association (GEE2)

In the last section we described estimating equations that allow a consistent estimation of the mean. We will now describe a set of equations that will allow the estimation of the first and the second moments jointly and consistently. These estimating equations are called GEE2. Note however, that Liang, Zeger and Quaqish (1992) used the term GEE2 only for the simultaneous estimation of the mean and the association. We believe that it is better to distinguish between estimating equation of first and second order.

Currently, no clear and unique definition of GEE2 is possible, as several procedures are summarised by this term. In this section, we will describe the development of the different estimating equation systems. The two systems, (3) and (7), have a comparable from and under certain regulation conditions the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ are asymptotically normal distributed.

The proof of the asymptotic distribution has been given by Prentice (1988) but without the exact definition of the regularity conditions. Prentice (1988) also gives the asymptotic covariance matrix. The two equations (3) and (7) can be imbedded into the Generalized Method of Moments (GMM) (see Hansen, 1982; Newey, 1993) as shown in (Ziegler, 1995) and therefore the asymptotic normality can be proven with the conditions given by Hansen (1982).

(3) and (7) can be summarised to one system of estimating equations:

$$\boldsymbol{u}\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \dfrac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'} & \boldsymbol{0} \\ \boldsymbol{0} & \dfrac{\partial \boldsymbol{\varrho}_i}{\partial \boldsymbol{\alpha}'} \end{pmatrix}' \begin{pmatrix} \mathbb{V}(\boldsymbol{y}_i) & \boldsymbol{0} \\ \boldsymbol{0} & \mathbb{V}(\boldsymbol{z}_i) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{y}_i - \boldsymbol{\mu}_i \\ \boldsymbol{z}_i - \boldsymbol{\varrho}_i \end{pmatrix} = \boldsymbol{0} \quad (8)$$

It can be seen that the matrix of the first derivatives and the working covariance matrix are matrices in a block-diagonal form. Therefore, (8) is a simplification of the following system:

$$\boldsymbol{u}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\beta}'} & \frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\alpha}'}\\ \frac{\partial\boldsymbol{\varrho}_i}{\partial\boldsymbol{\beta}'} & \frac{\partial\boldsymbol{\varrho}_i}{\partial\boldsymbol{\alpha}'}\end{pmatrix}'\begin{pmatrix}\mathbb{V}(\boldsymbol{y}_i) & \mathbb{C}ov(\boldsymbol{y}_i,\boldsymbol{z}_i)\\ \mathbb{C}ov(\boldsymbol{z}_i,\boldsymbol{y}_i) & \mathbb{V}(\boldsymbol{z}_i)\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{y}_i-\boldsymbol{\mu}_i\\ \boldsymbol{z}_i-\boldsymbol{\varrho}_i\end{pmatrix} = \mathbf{0} \quad (9)$$

Note that $\boldsymbol{\mu}_i$ is a function of the association parameter $\boldsymbol{\alpha}$, if $\partial\boldsymbol{\mu}_i/\partial\boldsymbol{\alpha}' \neq \mathbf{0}$.

However, this assumption is not always plausible. Additionally, it is difficult to interpret a mean vector that includes $\boldsymbol{\alpha}$. In most applications, the mean values are only defined as a function of $\boldsymbol{\beta}$—not depending on $\boldsymbol{\alpha}$. The form $\partial\boldsymbol{\mu}_i/\partial\boldsymbol{\alpha}' = \mathbf{0}$ in (9) implies also that the association—here the correlation—is a function of $\boldsymbol{\beta}$. In general, $\boldsymbol{\varrho}_i$ is defined with Fisher's z and therefore independent of $\boldsymbol{\beta}$. In most applications where the correlation was used, $\partial\boldsymbol{\varrho}_i/\partial\boldsymbol{\beta}' = \mathbf{0}$ was assumed.

If the matrix of the first derivatives has a block-diagonal form, then $\mathbb{V}$ has to be block diagonal, to guarantee unbiased estimators $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ (see Prentice and Zhao, 1991; Ziegler, 1994b). With this approach, modelling $\boldsymbol{\alpha}$ using Fisher's z yields (8), as $\boldsymbol{\beta}$ is not needed to model the association.

So far, we only defined estimating equations using the correlation as measurement for the association. However, the equations could also be defined using the covariance matrix. Then $s_{itt'} = (y_{it} - \mu_{it})(y_{it'} - \mu_{it'})$ and $\sigma_{itt'} = \mathbf{E}(s_{itt'}) = \mathbb{C}ov(y_{it}, y_{it'})$ are used instead of $z_{itt'}$ and $\varrho_{itt'}$. The first derivatives and the working variance matrices have to be changed accordingly.

The main question is, how to model the association between $\boldsymbol{\sigma}_i$ and $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}_i$ and $\boldsymbol{\beta}$ respectively. $\sigma_{itt'} = (v_{it}v_{it'})^{-1/2}\varrho_{itt'}$ and therefore $\sigma_{itt'}$ can be modelled as a function of $\boldsymbol{\beta}$ via v_{it} and as a function of $\boldsymbol{\alpha}$ via $\varrho_{itt'}$.

For this equation system the following holds: If $\boldsymbol{\mu}_i$ and $\boldsymbol{\sigma}_i$ are specified correctly as functions of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, then $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}$ can be estimated consistently and the estimators are asymptotically normal distributed (see Gourieroux and Monfort, 1984; Zhao and Prentice, 1990; Zhao and Prentice, 1991; Prentice and Zhao, 1991; Gourieroux and Monfort, 1993). The asymptotic covariance matrix is given e.g. by Prentice and Zhao (1991). The estimating equation system with the covariance matrix is not often used, probably due to the following disadvantage compared to (3) and (7): Here, it is necessary to specify $\boldsymbol{\mu}_i$ correctly as well as $\boldsymbol{\sigma}_i$ in order to obtain a consistent estimator of $\boldsymbol{\beta}$. Due to the independence of (3) and (7) and $\boldsymbol{\varrho}_i$, the estimator $\boldsymbol{\beta}$ is consistent even if $\boldsymbol{\varrho}_i(\boldsymbol{\alpha})$ is not specified correctly. The interpretation of the parameter $\boldsymbol{\alpha}$ is not improved by using $\boldsymbol{\sigma}_i$ instead of $\boldsymbol{\varrho}_i$. The advantage of using the covariance structure to model the association compared to the system (3) and (7) is that the estimation equations can be obtained using a (pseudo) maximum likelihood (PML2) approach (see Gourieroux and Monfort, 1984; Gourieroux and Monfort, 1993) as shown by (Ziegler, 1994b). This allows to define exact regularity conditions.

With the PML2 methods, the estimating equations can also be described using second moments (see Zhao and Prentice, 1990; Prentice and Zhao, 1991; Liang et al., 1992; Ziegler, 1994b). The relationship between the second moments and the log odds ratios can be described easily (Bishop, Fienberg and Holland, 1975). In this situation, the log odds ratio can be modelled as linear functions of the covariables $\boldsymbol{X}_i$ and the unknown parameter $\boldsymbol{\alpha}$.

If (9) is used together with the log odds ratio, then consistent estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ exist and are jointly (multivariate) normal distributed, if mean and association structure are specified correctly (see Zhao and Prentice, 1990; Prentice and Zhao, 1991; Gourieroux and Monfort, 1984; Gourieroux and Monfort, 1993; Ziegler, 1994b). The asymptotic covariance matrix is given in Liang et al. (1992). A misspecification of $\boldsymbol{\alpha}$ can lead to an inconsistent estimate of $\boldsymbol{\beta}$, since $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are estimated simultaneously.

The orthogonal models for the parameter $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (Liang et al., 1992) and the log odds ratio for the association structure can be used to transform the simultaneous estimation procedure of $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}$ into a two-step procedure. Then $\hat{\boldsymbol{\beta}}$ is consistent, even if $\boldsymbol{\alpha}$ is not specified correctly. This approach is called 'alternate logistic regression (ALR)' (Carey, Zeger and Diggle, 1993), here the logit-link is used as link function. The advantage of the two-step procedure was first observed by Firth, D. and Diggle, P. in their discussions of the paper by Liang et al. (1992). The ALR corresponds to the approach of Prentice (1988), except that the log odds ratio is used instead of the correlation. It can only be used if the response is categorical.

There is a further problem using marginal models for binary variables to estimate the mean and association structure: If the correlation or marginal log odds ratios were used instead of conditional log odds ratios, the parameter space of the association parameters is for correlation bounded if $T \geq 2$ and for marginal log odds ratios if $T \geq 3$ (see Fitzmaurice and Laird, 1993; Fitzmaurice, Laird and Lipsitz, 1994). These aspects are discussed in detail in Prentice (1988), Liang et al. (1992), Fitzmaurice and Laird (1993), Fitzmaurice et al. (1994), Ziegler (1994b) and Ziegler et al. (1996). A possible solution of this problem is to investigate the full likelihood (see Fitzmaurice and Laird, 1993; Fitzmaurice et al., 1994).

4 An example: A 2 × 2 Crossover Trial

The data given here to illustrate some practical issues are from a 2×2 crossover trial on cerebrovascular deficiency adapted from Jones and Kenward (1989), respectively Diggle, Liang and Zeger (1994). Treatments A and B are active drug and placebo, respectively. The outcome y_{it} indicates whether an electrocardiogram of the person i at time t, $t = 1, 2$ was judged abnormal ($y_{it} = 1$) or normal ($y_{it} = 0$). The results are presented in table 1.

	response			
Group	(1,1)	(0,1)	(1,0)	(0,0)
AB	22	0	6	6
BA	18	4	2	9

Table 1: response profiles 2×2 crossover trial

Among 34 persons that were treated first with the drug and then with the placebo, 28 had a normal result in the first period and 22 in the second period.

	Model		
Variable	1	2	3
constant	0.431 (1.209) [1.210]	0.666 (2.335) [2.313]	0.660 (2.056) [2.297]
period (x_1)	0.175 (0.347) [0.347]	-0.295 (-1.271) [-1.276]	-0.274 (-0.728) [-1.181]
treatment (x_2)	1.110 (1.934) [1.934]	0.669 (2.433) [2.444]	0.558 (1.475) [2.393]
interaction ($x_1 x_2$)	-1.023 (-1.045) [-1.045]	—	—
association α	1.447 (3.673) [3.216]	1.449 (3.705) [3.256]	—

Table 2: estimation results

Out of 33 persons with the combination BA, 20 normal values were observed in the first period and 22 in the second period.

We will use several GEE models for these data which include $n = 67$ observations, $T = 2$. The models include two covariables: $x_{i1} = 1$ if $t = 2$ and $x_{i1} = 0$, if $t = 1$, and $x_{i2} = 1$, if person i receives treatment A (Medicament), otherwise 0.

The link function is the logit link and for variance the binomial distribution is assumed. The association does not depend on the covariates. The estimators for the parameters of different models are given in table 2. Model-based z values are given in parentheses and robust z values are given in brackets. Note that our results are slightly different to those given by Diggle et al. (1994).

In this example, we used the GEE1 equations from Prentice (1988). The parameter for the association is defined by Fisher's z and the estimated correlations are the same for model 1 and model 2 ($\rho = 0.619$). Model 2 includes in contrast to model 1 not an interaction term and so the treatment effect becomes statistically significant ($p \leq 0.05$). Ignoring the association within the clusters (model 3), yields a non-significant treatment effect if the model-based variance is calculated. However, using the robust variance matrix yields a significant result ($p \leq 0.05$).

5 Recommendations for the Use of GEE

For practical use, some recommendations are needed to decide whether the ML methods for multivariate distributions (e.g. FGLS or the approach by Fitzmaurice and Laird (1993)) or GEE methods should be used.

In general, the ML method should only be used if the complete distribution of $\boldsymbol{y}_i$, conditional on $\boldsymbol{X}_i$, is specified correctly. If this is not the case, misspecification may yield inconsistent estimators of the parameters, either only for the asymptotic variance matrix or for both, the parameters and their asymptotic variance matrix. GEE1 yields inconsistent estimators for the mean, if it is not specified correctly. However, the association between observations within the clusters is treated as nuisance parameter. The use of the robust estimators for the variance has to be used, if misspecification of the association structure is possible. If the investigation of the association is the main goal of the analysis, GEE2 can be used, but only if the mean and association structure are specified correctly. GEE2 yields—if block-diagonal matrices are used—consistent estimation of the mean-structure even if the association is not specified correctly.

Several authors have investigated the efficiency and the consistency of the GEE1 approach: Paik (1988), Park (1993), Sharples (1989), Sharples and Breslow (1992), Lee, Scott and Soo (1993), McDonald (1993), Emrich and Piedmonte (1992), Royall (1986). However, the results are inconsistent and many questions remain open. The efficiency of GEE2 estimation has not been investigated in detail. Some theoretical results exist for the asymptotic distributions in the context of PML2 estimation (Gourieroux and Monfort, 1984) and for the GMM estimation (Newey, 1993).

Vach, Grömping and Schulz (1994) have shown that for panel data with time dependent exogenous covariables, IEE is preferable to avoid biases for $\boldsymbol{\beta}$, if the association structure is not specified correctly. For a detailed discussion see also Sullivan Pepe and Anderson (1994)

Lee et al. (1993) have shown for a simple model that the robust variance estimation for $\hat{\boldsymbol{\beta}}_{IEE}$ yields an estimator that underestimates the true covariance matrix. However, all estimation procedures—including ML—yield an underestimation of the covariance matrix. As the covariance can be estimated consistent, small sample size yields biased estimation. This bias decreases with the number of clusters n (see Sharples, 1989; Sharples and Breslow, 1992).

It was noted that GEE2 algorithm does converge less often then GEE1. To apply GEE2, a simple structure of the working matrix is recommended. If convergence problems occur, it is recommended to use e.g. the idendity matrix as upper right block of the working matrix. Further simplification is obtained by setting the third moments to $\mathbf{0}$. This working matrix is called "working-covariance matrix for applications" (Kastner, 1994).

From published theoretical results and our own experience we recommend that GEE should only be used if at least 30 clusters with $T \leq 4$ are available. Liang and Zeger (1986) have proposed the GEE as a more efficient procedure than IEE but the authors found in their own application that only a small gain in efficiency was obtained using the working correlation matrix. We recommend to use IEE first and to model other association structures in a second step.

The problem of missing values for the GEE has been investigated recently by Ziegler (1994b), Ziegler (1996) and in Robins, Rotnitzky and Zhao (1995), Robins and Rotnitzky (1995), Rotnitzky and Robins (1995).

A detailed description of the test problem with GEE and in connection to the pseudo maximum mikelihood methods is given in Rotnitzky (1988), Rotnitzky and Jewell (1990), Gourieroux and Monfort (1993), Arminger (1995) and Ziegler (1996). Also, regression diagnostic techniques for the GEE have been well-developed (see Hall, Zeger and Bandeen-Roche, 1994; Ziegler, Bachleitner and Arminger, 1995).

An extension of the univariate GLM is the multivariate GLM. While in the univariate GLM a parameter vector $\boldsymbol{\beta}$ is used that is the same for all i, this is no longer necessary in the multivariate GLM. GEE can be extended in an analogous manner so that the parameter vectors $\boldsymbol{\beta}_t$ can be estimated. This extension is important for longitudinal studies where the influence of the covariates changes with time. This extension is well-described in several papers, e.g. by Wei and

Stram (1988), Lipsitz, Kim and Zhao (1994), Ziegler and Arminger (1995) or Ziegler (1994b).

Ordered categorical and non-ordered categorical data are discussed in some papers mainly as in this context convergence problems occur frequently (see Stram, Wei and Ware, 1988; Lipsitz, Kim and Zhao, 1994; Clayton, 1992; Ziegler, 1994a; Ziegler, 1994b; Miller, 1995; Miller, Davis and Landis, 1993; Liang et al., 1992; Kenward, Lesaffre and Molenberghs, 1994; Kastner, 1994).

Several programs are available for the application of GEE (see Karim and Zeger, 1988; Lipsitz and Harrington, 1990; Davis, 1993; Grömping, 1993; Kastner, 1994; Ziegler, 1994b), however, the robust variance matrix can also be estimated using Jack-knife techniques (Lipsitz, Dear and Zhao, 1994).

Acknowledgements

C. Kastner is supported by the Deutsche Forschungsgemeinschaft, SFB 386, Teilprojekt C3.

References

Arminger, G. (1995). Specification and estimation of mean structures, *in* G. Arminger, C. C. Clogg and M. E. Sobel (eds), *Handbook of Statistical Modeling for the Behavioral Sciences*, Plenum, New York.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*, MIT Press, Cambridge.

Breslow, N. (1990). Tests of hypotheses in overdispersed poisson regression and other quasi–likelihood models, *Journal of the American Statistical Association* **85**: 565–571.

Carey, V., Zeger, S. L. and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regression, *Biometrika* **80**: 517–526.

Cessie, S. l. and Houwelingen, J. v. (1994). Logistic regression for correlated binary data, *Applied Statistics* **43**: 95–108.

Clayton, D. (1992). Repeated ordinal measurements: A generalised estimating equation approach, *to be published* .

Cochrane, D. and Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms, *J. Amer. Statist. Assoc.* **44**: 32–61.

Crowder, M. (1985). Gaussian estimation for correlated binary data, *Journal of the Royal Statistical Society, Series B* **47**: 229–237.

Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measurements, *Biometrika* **82**: 407–410.

Davis, C. S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials, *Stat. in Med.* **10**: 1959–1980.

Davis, C. S. (1993). A computer program for regression analysis of repeated measures using generalized estimating equations, *Computer Methods and Programs in Biomedicine* **40**: 15–31.

Diggle, P. J., Liang, K.-Y. and Zeger, S. L. (1994). *Analysis of longitudinal data*, Oxford University Press, New York.

Emrich, L. J. and Piedmonte, M. R. (1992). On some small sample properties of generalized estimating equation estimates for multivariate dichotomous outcomes, *Journal of statistical computation and simulation* **41**: 19–29.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models*, Springer, New York.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses, *Biometrika* **80**: 141–151.

Fitzmaurice, G. M., Laird, N. M. and Lipsitz, S. R. (1994). Analysing incomplete longitudinal binary responses: a likelihood-based approach, *Biometrics* **50**: 601–612.

Fitzmaurice, G. M., Laird, N. M. and Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses, *Statistical Science* **8**: 284–309.

Glynn, R. J. and Rosner, B. (1994). Comparison of alternative regression models for paired binary data, *Statistics in Medicine* **13**: 1023–1036.

Gourieroux, C. and Monfort, A. (1984). Pseudo maximum likelihood methods: Theory, *Econometrica* **52**: 682–700.

Gourieroux, C. and Monfort, A. (1993). Pseudo-likelihood methods, *in* G. Maddala, C. R. Rao and H. Vinod (eds), *Handbook of Statistics*, Vol. 11, Elvesier, Amsterdam.

Greene, W. H. (1993). *Econometric Analysis*, 2 edn, Mac Millan, New York.

Grömping, U. (1993). GEE: A sas macro for longitudinal data analysis, *Technical report*, Universität Dortmund.

Hall, C., Zeger, S. L. and Bandeen-Roche, K. (1994). Added variable plots for regression with dependent data, *Technical report*, Department of Biostatistics, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD.

Hansen, L. (1982). Large sample properties of generalized methods of moments estimators, *Econometrica* **50**: 1029–1055.

Jones, B. and Kenward, M. G. (1989). *Design and Analysis of Crossover Trials*, Chapman and Hall, London.

Karim, M. and Zeger, S. L. (1988). GEE: A sas macro for longitudinal analysis, *Technical report*, Department of Biostatistics, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD.

Kastner, C. (1994). *Marginale Modelle für korrelierte diskrete Variable*, Master's thesis, Institut für Statistik der Universität München.

Kenward, M. G., Lesaffre, E. and Molenberghs, G. (1994). Application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random, *Biometrics* **50**: 945–953.

Lee, A., Scott, A. and Soo, S. (1993). Comparing liang–zeger estimates with maximum likelihood in bivariate logistic regression, *Statistical Computation and Simulation* **44**: 133–148.

Lefkopoulou, M., Moore, D. and Ryan, L. (1989). The analysis of multiple correlated binary outcomes: Application to rodent teratology experiments, *JASA* **84**: 810–815.

Lefkopoulou, M. and Ryan, L. (1993). Global tests for multiple binary outcomes, *Biometrics* **49**: 975–988.

Liang, K.-Y. (1992). Extensions of generalized linear models in the past twenty years: Overview and some biomedical applicationsglobal tests for multiple binary outcomes, 16*th* International Biometric Conference, Hamilton, New Zealand.

Liang, K.-Y. and Beaty, T. (1991). Measuring familial aggregation by using odds ratio regression models, *Genetic Epidemiology* **8**: 361–370.

Liang, K.-Y. and Hanfelt, J. (1992). Quasi–likelihood and robust variance estimate, *Technical report*, Department of Biostatistics, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD.

Liang, K.-Y. and Hanfelt, J. (1994). On the use of the quasi–likelihood method in teratological experiments, *Biometrics* **50**: 872–880.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.

Liang, K.-Y. and Zeger, S. L. (1993). Regression analysis for correlated data, *Annu. Rev. Pub. Health* **14**: 43–68.

Liang, K.-Y., Zeger, S. L. and Quaqish, B. (1992). Multivariate regression analysis for categorical data, *J. Royal. Statist. Soc. B* **54**: 3–40.

Lipsitz, S. R., Dear, K. B. and Zhao, L. P. (1994). Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data, *Biometrics* **50**: 842–846.

Lipsitz, S. R., Fitzmaurice, G., Orav, E. and Laird, N. M. (1994). Performance of generalized estimating equations in practical situations, *Biometrics* **50**: 270–278.

Lipsitz, S. R. and Harrington, D. P. (1990). Analyzing correlated binary data using sas, *Computers and Biomedical Research* **23**: 268–282.

Lipsitz, S. R., Kim, K. and Zhao, L. P. (1994). Analysis of repeated categorical data using generalized estimating equations, *Statistics in Medicine* **13**: 1149–1163.

Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association, *Biometrika* **78**: 153–160.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chapmann and Hall, London.

McDonald, B. W. (1993). Estimating logistic regression parameters for bivariate binary data, *J. Royal. Statist. Soc. B* **55**: 391–397.

Miller, M. E. (1995). Analysing categorical responses obtained from large clusters, *Applied Statistics* **44**: 173–186.

Miller, M. E., Davis, S. C. and Landis, J. R. (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares, *Biometrics* **49**: 1033–1044.

Muñoz, A., Schouten, J., Segal, M. and Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data, *Biometrika* **48**: 733–742.

Nelder, J. and Wedderburn, R. W. M. (1972). Generalized linear models, *J. Royal Statist. Soc. A* **135**: 370–384.

Neuhaus, J., Kalbfleisch, J. and Hauck, W. (1991). A comparison of cluster–specific and population–averaged approaches for analyzing correlated binary data, *International Statistical Review* **59**: 25–35.

Newey, W. K. (1990). Semiparametric efficiency bounds, *Journal of Applied Econometrics* **5**: 99–135.

Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions, *in* G. Maddala, C. R. Rao and H. Vinod (eds), *Handbook of Statistics*, Vol. 11, Elvesier, Amsterdam.

Paik, M. C. (1988). Repeated measurement analysis for nonnormal data in small samples, *Communications in Statistics and Simulation, Series B* **17**: 1155–1171.

Paik, M. C. (1992). Parametric variance function estimation for nonnormal repeated measurement data, *Biometrics* **48**: 19–30.

Park, T. (1993). A comparison of the generalized estimating equation approach with the maximum likelihood approach for repeated measurements, *Statistics in Medicine* **12**: 1723–1732.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation, *Biomerics* **44**: 1033–1048.

Prentice, R. L. and Zhao, L. P. (1991). Estimation equations for parameters in means and covariance of multivariate discrete and continous responses, *Biomerics* **47**: 825–839.

Qaqish, B. and Liang, K.-Y. (1992). Marginal models for correlated binary response with multiple classes an multiple levels of nesting, *Biometrics* **49**: 939–950.

Qin, L. and Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters, *Canadian Journal of Statistics* **23**: 145–159.

Robins, J. M. and Rotnitzky (1995). Semiparametric efficiency in multivariate regression models with missing data, *J. Amer. Statist. Assoc.* **90**: 122–129.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models repeated outcomes in the presence of missing data, *Journal of the American Statistical Association* **90**: 106–120.

Rotnitzky, A. (1988). *Analysis of Generalized Linear Models for Cluster Correlated Data*, PhD thesis, University of California, Berkeley.

Rotnitzky, A. and Jewell, N. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data, *Biometrika* **77**: 485–497.

Rotnitzky, A. and Robins, J. M. (1995). Semiparametric estimation of models for means and covariances in the presence of missing data, *Scandinavian Journal of Statistics* **22**: 323–333.

Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimation, *International Statistical Review* **54**: 221–226.

Sharples, K. (1989). *Regression Analysis of Correlated Binary Data*, PhD thesis, University of Washington.

Sharples, K. and Breslow, N. (1992). Regression analysis of correlated binary data: Some small sample results for the estimating equation approach, *Statistical Computation and Simulation* **42**: 1–20.

Stram, D., Wei, L. and Ware, J. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations and time–dependent covariates, *Journal of the American Statistical Association* **83**: 631–637.

Sullivan Pepe, M. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data, *Communications in Statistics—Simulation and Computation* **23**(4): 939–951.

Thall, P. and Vail, S. (1990). Some covariance models for longitudinal count data with overdispersion, *Biometrics* **46**: 657–671.

Vach, W., Grömping, U. and Schulz, K. (1994). On the relation between parameter estimates from transition and marginal models, *submitted to Biometrics* .

Wei, L. and Stram, D. (1988). Analysing repeated measurements with possibly missing observations by modelling marginal distributions, *Statistics in Medicine* **7**: 139–148.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models, *Journal of the American Statistical Association* **76**: 419–433.

White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica* **50**: 1–25.

Zeger, S. L. (1988). Commentary, *Stat.in Med.* **7**: 95–107.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes, *Biometrics* **42**: 121–130.

Zeger, S. L., Liang, K.-Y. and Self, S. G. (1985). The analysis of binary longitudinal data with time-independent covariates, *Biometrika* **72**: 31–38.

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a generalized quadratic model, *Biometrika* **77**: 642–648.

Zhao, L. P. and Prentice, R. L. (1991). Use of a quadratic exponential model to generate estimating equations for means, variances, and covariances, *in* V. P. Godambe (ed.), *Estimating Functions*, University Press, Oxford.

Zhao, L. P., Prentice, R. L. and Self, S. G. (1992). Multivariate mean parameter estimation by using a partly exponential model, *JRSS B* **54**: 805–811.

Ziegler, A. (1994a). GEE1: Ein programmsystem zur schätzung von parameterstrukturen in multivariaten verallgemeinerten linearen modellen mit generalized estimating equations, *Technical report*, Fachbereich Wirtschaftswissenschaft, BUGH Wuppertal.

Ziegler, A. (1994b). *Verallgemeinerte Schätzgleichungen zur Analyse korrelierter Daten*, PhD thesis, Universität Dortmund, Fach Statistik.

Ziegler, A. (1995). The different parameterizations of the GEE1 and the GEE2, *in* G. U. H. Seeber, B. Francis, R. Hatzinger and G. Steckel-Berger (eds), *Statistical Modelling*, Vol. 104 of *Lecture Notes in Statistics*, Springer, New York.

Ziegler, A. (1996). Pseudo maximum likelihood estimation and generalized estimating equations, *to be published* .

Ziegler, A. and Arminger, G. (1995). Analyzing the employment status with panel data from the gsoep – a comparison of the mecosa and the gee1 approach for marginal models, *Vierteljahreshefte zur Wirtschaftsforschung* **64**: 72–80.

Ziegler, A., Bachleitner, R. and Arminger, G. (1995). Pseudo maximum likelihood schätzung und regressionsdiagnostik für zähldaten: Suizid durch tourismus, *Allgemeines Statistisches Archiv* **79**: 170–195.

Ziegler, A., Kastner, C., Grömping, U. and Blettner, M. (1996). Generalized estimating equations: Herleitung und anwendung, *Informatik, Epidemiologie und Biometrie in Medizin und Biologie* **27**(2): 111–?