Chapter 2

GEE for Longitudinal Ordinal Data: Comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN

N. Nooraee, G. Molenberghs, and E.R. van den Heuvel

Computational Statistics & Data Analysis. 2014; 77: 70-83.

2.1 Abstract

Studies in epidemiology and social sciences are often longitudinal and outcome measures are frequently obtained by questionnaires in ordinal scales. To understand the relationship between explanatory variables and outcome measures, generalized estimating equations can be applied to provide a population-averaged interpretation and address the correlation between outcome measures. It can be performed by different software packages, but a motivating example showed differences in the output. This paper investigated the performance of GEE in R (version 3.0.2), SAS (version 9.4), and SPSS (version 22.0.0) using simulated data under default settings. Multivariate logistic distributions were used in the simulation to generate correlated ordinal data. The simulation study demonstrated substantial bias in the parameter estimates and numerical issues for data sets with relative small number of subjects. The unstructured working association matrix requires larger numbers of subjects than the independence and exchangeable working association matrices to reduce the bias and diminish numerical issues. The coverage probabilities of the confidence intervals for fixed parameters were satisfactory for the independence and exchangeable working association matrix, but they were frequently liberal for the unstructured option. Based on the performance and the available options, SPSS and multgee, and repolr in R all perform quite well for relatively large sample sizes (e.g. 300 subjects), but multgee seems to do a little better than SPSS and repolr in most settings.

Key words: Correlated ordinal data, generalized estimating equations, copula, multivariate logistic distribution, Bridge distribution.

2.2 Introduction

2.2.1 Motivating example

Change in Quality of Life was investigated in a study of women who underwent a laparoscopic hysterectomy (surgery). In total, 72 patients were measured using the Short Form-36 Health Survey questionnaire before surgery (baseline), and six weeks after surgery, and then six months after surgery. One specific domain is the emotional role (ER). It was scored with just one item having four possible outcome levels, coded $\{1, 2, 3, 4\}$. Higher scores indicate a higher quality of life. The goal was to investigate whether ER was affected by surgery and to determine the role of some explanatory variables, such as age (a), comorbidity (cm), blood loss (bl) and complications (c) during surgery, and duration (d) of surgery. We decided to implement the following model

$$\begin{aligned} &logit\left[\mathbb{P}\left(O_{ij} \leq c\right)\right] \\ &\qquad = \beta_{0c} + \beta_1 a_i + \beta_2 cm_i + \beta_3 t_{ij} + \left(\alpha_1 bl_i + \alpha_2 c_i + \alpha_3 d_i\right)\delta_{t_{ij}}, \end{aligned}$$

with O_{ij} the j^{th} ordinal response for subject i, t_{ij} the j^{th} time moment for subject i ($t_{i1} = 0$, $t_{i2} = 6$, and $t_{i3} = 26$ weeks), and with δ_x an indicator variable equal to one when $x > 0$ and zero otherwise. The indicator δ_x is needed because the covariates bl, d, c can only affect ER after surgery. The parameter β_3 would indicate the effect of surgery over time when corrected for other variables.

We decided to estimate the parameters with generalized estimating equations (GEE) to obtain a population-averaged interpretation and to address the correlation between subject outcomes. We applied the geepack (ordgee function), repolr (repolr function) and multgee (ordLORgee function) packages in R under default settings and selected the most complex working association structure available in each package: unstructured working association in geepack and multgee, and exchangeable working correlation in repolr. Geepack and multgee provided surprisingly different results (Table 2.1), while repolr produced no parameter estimates due to the estimation of cell probabilities equal to one. The highest score of 4 was indeed frequently observed: almost 90 percent after six months of surgery. Not yet completely satisfied with the results, we decided to analyse this data also with SAS (GENMOD procedure) and SPSS (GENLIN command) to verify the parameter estimates of multgee and geepack. We chose unstructured working correlation ma-

trix in SPSS and independence structure in SAS, based on the options available. Similar to repolr, SPSS did not converge, but SPSS was able to produced results with the exchangeable correlation structure. The results are listed again in Table 2.1.

Table 2.1 The parameter estimates (robust/empirical standard error) under an independent working correlation matrix.

Parameters	geepack Unstructured	multgee Unstructured	SPSS Exchangeable	SAS Independent	multgee Exchangeable
Threshold 1	0.702(1.040)	0.533(0.947)	0.846(0.932)	0.289(1.063)	0.647(0.995)
Threshold 2	1.139(1.020)	1.007(0.907)	1.306(0.902)	0.739(1.009)	1.090(0.918)
Threshold 3	1.464(1.019)	1.368(0.925)	1.657(0.916)	1.077(1.023)	1.438(0.934)
Age	-0.035(0.019)	-0.033(0.017)	0.039(0.017)	-0.026(0.019)	-0.034(0.017)
Comorbidity	-0.910(0.551)	-0.506(0.411)	0.508(0.414)	-0.631(0.439)	-0.556(0.421)
Time	-0.025(0.020)	-0.025(0.017)	0.025(0.018)	-0.022(0.018)	-0.023(0.017)
Blood loss	-0.002(0.001)	-0.002(0.001)	0.003(0.001)	-0.002(0.001)	-0.003(0.001)
Complication	0.592(1.001)	1.832(0.677)	-1.725(0.626)	1.812(0.800)	1.700(0.654)
Duration	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)

Comparing the results demonstrates several differences. First, not all packages seem to converge, but secondly, there exist differences in the parameter estimates between the packages (Table 2.1). This could be due to the different choices in correlation structure, but differences remain even when the same class of structure is chosen. Indeed, as we already mentioned, geepack and multgee provided different results for the unstructured association, but also SPSS and multgee produce different results under the exchangeable structure (Table 2.1). Not only did the estimates differ in this case, they are also opposite in sign. When each package is run with the independence structure, all packages are identical (to the results of SAS in Table 2.1), except for geepack, which leads to completely different results, and for SPSS, which gives opposite signs, but the same absolute numbers.

These different results in performance and in estimates encouraged us to investigate the similarities and discrepancy between the GEE methods in R (version 3.0.2), SAS (version 9.4), and SPSS (version 22.0.0) for longitudinal ordinal data using simulation studies. In these studies we would know what mean models the software should estimate. Note that they all estimate the same mean model, and that they treat the associations as nuisance parameters, although they may have implemented different association structures (even in the same class).

2.2.2 Background

Generalized estimating equations (GEE) were introduced by Liang and Zeger[18,52] as general approach for handling correlated discrete and continuous outcome variables. It only requires specification of the first moments, the second moments, and correlation among the outcome variables. The goal of this procedure is to estimate fixed parameters without specifying the joint distribution. Prentice[39] extended the GEE approach by improving the estimation of the correlation parameters using a second set of equations based on Pearson's residuals, see also Lipsitz and Fitzmaurice[20]. Others modeled the association parameter as an odds ratio[22,19,4]. An alternative approach considered latent variables with a bivariate normal distribution underneath the correlated binary variables, see Qu et al.[40].

Extending GEE to ordinal data is not immediately obvious because the first and second moments are not defined for ordinal observations. It requires the introduction of a vector of binary variables that relates one-to-one to the ordinal variables[7]. With this set of binary variables the original GEE method[18,52] as well as the method for estimation of the association parameters can be extended to ordinal data[21,11,38,44]. Different approaches have been used to estimate the association parameters in GEE. Lipsitz et al.[21] used Pearson's residual, while Parsons et al.[38] minimized the logarithm of the determinant of the covariance matrix of the fixed parameters, i.e. minimized the standard errors of the parameter estimates.

Instead of using correlations, Lumley[24] applied common odds ratios for the association of multivariate ordinal variables to reduce the number of association parameters. Williamson et al.[46] suggested a GEE method for bivariate ordinal responses with the global odds ratio as measure of dependency. In this context, two sets of equations were used: one for the fixed parameters and one for the association parameters. To make the approach available to others, Williamson et al.[47] developed two SAS macros but they were not officially incorporated in SAS. Yu and Yuan[51] developed one macro that extended these two macros to unbalanced data and it is only available upon request from the authors. The approach with two sets of equations was further extended to multivariate ordinal outcomes using global odds ratios as measure of dependency, while the two sets of equations can be integrated into one set of equations for the fixed and association parameters simultaneously (see Heagerty and Zeger[11]). Nores and del Pilar Díaz[31]

investigated the efficiency and convergence of this approach via simulation. They applied function ordgee of R. Recently, Touloumis et al.[44] extended the GEE method for ordinal outcomes by considering local odds ratios as the measure of association.

Several overviews of GEE have been provided. Ziegler et al.[55] developed a bibliography of GEE, and Zorn[56] indicated the use of GEE in Political science. To recent books of of Ziegler[53]; Hardin and Hilbe[53,10] were fully dedicated to GEE, while Agresti and Natarajan[2], Liu and Agresti[23], and Agresti[1] discussed comprehensive reviews of more general models and methods for (correlated) categorical data. Two particular overviews focused on the models and tests that were programmed in the software packages LogXact 4.1, SAS 8.2, Stata 7, StatXact 5, and Testimate 6 for (correlated) categorical outcomes, including GEE[32,33]. Oster and Hilbe[34] also presented a general overview of software packages on exact methods, but they did not investigate the performance of these packages. Ziegler and Gromping[54]; Horton and Lipsitz[13] did compare software packages for the analysis of correlated data via GEE, but they focused on binary outcomes only. A comprehensive comparison of frequently used software packages for correlated ordinal data using GEE has not yet been conducted.

We applied a simulation study to compare the functions ordgee in geepack, ordLORgee in multgee and repolr in package repolr in R 3.0.2, the procedure GENMOD in SAS 9.4, and finally the procedure GENLIN in SPSS 22.0.0. We took the perspective of a general user with limited knowledge of the mathematical and numerical details of GEE. This means that we mainly used default settings in the simulation study. We simulated moderately to highly correlated multivariate logistic distributed latent variables using copula functions to obtain correlated ordinal data. This choice implies the logit models for the marginal distributions, but the correlation between the binary variables coding the ordinal outcomes is different from choices implemented in the software. We investigated the frequency of simulation runs with numerical convergence issues, and the bias in parameter estimates. We reported the coverage probabilities of the confidence intervals on these parameters using the Wald statistic. Finally we provided rejection rates of the proportionality test (if available).

2.3 Generalized estimating equations

Generalized estimating equations for ordinal outcomes require several aspects. The first aspect is to choose a model for the covariates and a non-linear link function to connect the model to the cumulative probabilities. Then the second aspect is to create a set of binary variables describing all possible outcomes for the ordinal observations[7]. The third aspect is to choose a working correlation matrix or working association structure to describe the possible association between all binary variables. The fourth and final aspect is the estimation method for the association parameters involved in the association structure.

To illustrate these aspects in more detail, consider a random sample of observations from n subjects. Let $O_i = \left(O_{i1}, O_{i2}, ..., O_{in_i}\right)$ be the ordinal responses of subject i and O_{it} takes values in $\{1, 2, ..., C\}$ and let ${X_i}^\top = \left(X_{i1}^\top, X_{i2}^\top, ..., X_{in_i}^\top\right)$ be a $p \times n_i$ dimensional matrix of time varying and/or time stationary covariates for subject i. Then the connection between the covariates and the conditional probabilities of each ordinal outcome is described by

$$h[\mathbb{P}(O_{it} \leq c | X_{it} = x_{it})] = \beta_{0c} + x_{it}^\top \beta_1, \tag{2.1}$$

for $c = 1, 2, ..., C-1$, β_{0c} the threshold parameter for level c, β_1 the vector of regression coefficients corresponding to the covariates and with h a known link function. Any monotone increasing function h which would transfer the interval $(0, 1)$ to $(-\infty, \infty)$ could be applied as the link function[26], e.g. logit, probit and complementary log-log. The cumulative logits model is very popular for clustered ordinal outcomes due to its simple and comprehensive interpretation, the same as in logistic regression. This model is often referred to as the proportional odds model[1]. The cumulative probabilities with probit link function is more popular in econometrics, but then the model should no longer be interpreted as an odds ratio. The formulation in (2.1) is ascending in terms of level of ordinal outcomes but the model can be changed to descending in which $O_{it} \leq c$ is replaced by $O_{it} > c$.

There are three options for choosing the binary variables $Y_{it}^\top$ =(Y_{it1}, Y_{it2}, ..., Y_{itC-1}), with dimension $C-1$. The first option selects $Y_{itc} = I(O_{it} = c)$ (see Lipsitz et al.[21], and Touloumis et al.[44]) the second option selects $Y_{itc} = I(O_{it} > c)$ (see Heagerty and Zeger[11]), and finally the third option selects $Y_{itc} = I(O_{it} \leq c)$ (see Parsons et al.[38]). Note that for all options $c = 1, 2, ..., C-1$ and $I(.)$ is the indicator function equal to one when the argument is true and zero otherwise.

Consequently, the mean vector $\mu_i = \mathbb{E}(Y_i|X_i = x_i)$ is the mean of all binary variables $Y_i^\top = (Y_{i1}^\top, ..., Y_{in_i}^\top)$. Now the vector of regression parameter $\beta = (\beta_{01}, \beta_{02}, ..., \beta_{0C-1}, \beta_{11}, \beta_{12}, ..., \beta_{1p})^\top$ can be estimated using the GEE method by solving

$$u(\beta) = \sum_{i=1}^{N} D_i^\top V_i^{-1} [Y_i - \mu_i] = 0, \tag{2.2}$$

where $D_i = \partial\mu_i/\partial\beta$ and V_i is the so-called weight matrix or working covariance matrix of $\mathbf{Y}_i$. This matrix may depend on the vector of parameters β and the vector of association parameters α for the binary variables.

Liang and Zeger[18]; Lipsitz et al.[21] showed that given any parameterisations of the matrix V_i and assuming that the marginal model (2.1) is correctly specified, the solution $\widehat{\beta}$ for (2.2) is a consistent estimator of β and $\sqrt{n}(\widehat{\beta} - \beta)$ has an asymptotic multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $V_\beta = \lim_{n\to\infty} nV_\beta(n)$, with $V_\beta(n)$ defined by

$$V_\beta(n) = \Big(\sum_{i=1}^{n} D_i^\top V_i^{-1} D_i\Big)^{-1} \Big[\sum_{i=1}^{n} D_i^\top V_i^{-1} \mathsf{COV}\Big(Y_i) V_i^{-1} D_i\Big] \Big(\sum_{i=1}^{n} D_i^\top V_i^{-1} D_i\Big)^{-1}. \tag{2.3}$$

This form of variance is referred to as the empirical or robust variance estimator since it provides a consistent estimator regardless of the (mis)specification of V_i[21]. A model-based standard error would be obtained when $\mathsf{COV}\,(Y_i)$ in (2.3) is replaced by matrix V_i and then the covariance matrix in (2.3) would reduce to the last term in (2.3), which means that $V_\beta^{-1}(n) = \sum_{i=1}^{n} D_i^\top V_i^{-1} D_i$. It should be noted however, that the choice for a model-based estimator does not imply that the working covariance matrix V_i for the binary vector Y_i is a true covariance matrix. Issues related to covariance matrices for multivariate binary outcome variables were discussed by[5,6]. Fortunately, these issues do not cause difficulties in applying GEE, since the multivariate distribution can always partially be described by semi-parametric models, see Molenberghs and Kenward[28].

To be able to determine GEE estimates, the vector of association parameters α should be estimated. Commonly, the matrix V_i is re-parameterized by

$$V_i = A_i^{\frac{1}{2}} \mathsf{R}_i(\alpha) A_i^{\frac{1}{2}}, \tag{2.4}$$

with A_i a $n_i(C-1) \times n_i(C-1)$ diagonal matrix with elements given by the variance of the binary variable Y_{itc}, and the matrix $\mathsf{R}_i(\alpha)$ consists of the associations

between the binary variables. The R_i matrix contains three parts of associations. The first part is the association between the binary variables at one time point. The second one is the association of the same coded binary variables across time, and the third and final part is the association of two differently coded binary variables across time. Thus the "variance" of each ordinal outcome and the association between any pair of ordinal outcomes are represented by matrices rather than scalers.

Although Pearson's correlation has been applied to the association between binary variables within the same time point, different association measures have been applied to model the dependency between binary variables across time. Lipsitz et al.[21] assumed Pearson's correlation for all associations between binary variables and estimate the association parameters α with Pearson's residuals. Restricting to the logit link function, Parsons et al.[38] described the association between each pair of the binary variables over time as a product of a function of single parameter α and Pearson's correlation of the same pair of binary variable within a time point, i.e. $g_{st}(\alpha)\exp(-|\beta_{0c}-\beta_{0k}|/2)$, see Kenward et al.[15]. This scaler parameter is estimated by minimizing the logarithm of the determinant of the covariance matrix $(\log|\widehat{V}_\beta(n)|)$ of the parameter estimates in each step of the fitting algorithm for solving (2.2). As an alternative to Pearson's correlation, one can use the odds ratio. Heagerty and Zeger[11] applied global odds ratios for the association of repeated binary variables in the matrix V_i. They applied a second set of estimating equations of the form (2.2) to obtain these association parameters. This choice was first introduced for binary outcomes by Prentice[39], and for ordinal outcomes by Miller et al.[27]. Touloumis et al.[44] utilized local odds ratios to capture the association parameters in the V_i matrix. They used the Goodman's row and column effects model[9] to reparameterize the local odds, and then estimated the parameters using the iterative proportional fitting procedure[8].

All papers use the same definition for an independence, exchangeable and unstructured association matrix, but this does not imply that they fit identical associations. Under the independence working assumption all off-diagonal blocks are constant and equal to zero. Exchangeability over time indicates that the association between the binary variables Y_{itc} and Y_{isk}, for time t and s, $t \neq s$, is independent of time, but it may depend on the levels c and k. Finally, for unstructured associations there are no restrictions implied.

2.3.1 Software packages

R

R software offers several options for fitting marginal ordinal models with GEE approach. In the current paper, we compare all three packages: geepack, repolr and multgee.

The function ordgee in geepack[49] produces estimations according to the method of Heagerty and Zeger[12]. The function allows to choose logit, probit, and complementary log-log link function as well as four association structures (independence, exchangeable, unstructured, and user-defined). Selection of initial values for the regression parameters and odds ratio parameters are possible. An option is also available to change the default setting from descending to ascending. This package offers the robust estimator for the covariance matrix of the fixed parameters and the Wald statistic for testing the statistical significance of each coefficient in the model. The numerical procedure is the Fisher-scoring algorithm with the default number for the maximum number of iterations 25. The iteration procedure stops when the change in the parameter estimates is less than 0.0001.

Another function, repolr, is implemented in the repolr package[36]. The gee package needs to be uploaded in advance. The function repolr has implemented the GEE method of Parsons et al.[37] which exclusively supports the logit link function. An independence, uniform (exchangeable) and AR(1) working correlation structure can be selected. Under the uniform (exchangeable) assumption each element on the off-diagonal block of the matrix $\mathsf{R}_i(\alpha)$ is defined with $\mathsf{CORR}(Y_{itc}, Y_{isk}) = \alpha \exp(-|\beta_{0c} - \beta_{0k}|/2)$. Under AR(1) the correlation is $\mathsf{CORR}(Y_{itc}, Y_{isk}) = \alpha^{|s-t|}\exp(-|\beta_{0c} - \beta_{0k}|/2)$. Repolr has an option to choose an initial value for the correlation parameter. It automatically provides standard errors of the parameter estimates based on both the robust estimator and the model-based estimator. Furthermore, the function also has an option to test the proportionality assumption based on the score test[41]. The numerical procedure is the Newton-Raphson algorithm and the default setting for convergence of the numerical procedure is a relative change in parameters estimates less than 0.001 or a maximum number of iterations equal to 10.

The final option to peform GEE for ordinal outcomes in R is to use the ordLORgee function which is implemented in the multgee package[43]. This package requires the user to upload gnm and VGAM packages in advance. The ordLORgee

function uses the method of Touloumis et al.[44]. Beside the cumulative link functions logit, probit, cauchit, and cloglog, it could also fit adjacent-category logit models. An independence, uniform, category exchangeable, time exchangeable, unstructured, and a user-defined structure can be used. Under the uniform structure, all local odds ratios are identical. Relaxing this structure a little bit leads to two types of exchangeability assumptions. The category exchangeability structure assumes that local odds ratios are the same within time pairs, but could still be different between time pairs. An alternative is time exchangeability, which assumes that local odds ratios are independent of time, but could change with ordinal levels. Finally, the unstructured association is indicated by RC in this package. Robust and model-based variance estimators can be used. The numerical procedure for estimation of the fixed effects uses the Fisher-scoring algorithm. It stops if the relative change in the parameter estimates between two successive iterations is smaller than 0.001, or whenever it completes to 15 iterations.

SAS

PROC GENMOD in SAS software is a procedure to fit models for correlated binary and ordinal data (see Stokes et al.[42]). The SAS system selects $Y_{itc} = I(O_{it} = c)$ as binary coding for the ordinal outcome. It can also change from ascending to descending. The procedure supports logit, probit, and complementary log-log link function, but is limited to the independence working correlation matrix for ordinal data. It is possible to specify initial values for the regression parameters. The quasi information criterion (QIC)[35], type I and type III testing using either the Wald statistic or the generalized score statistic are available in this procedure.

Although the default setting in SAS is the use of the robust estimator, model based standard errors can be obtained as well. The numerical procedure for solving (2.2) is Newton-Raphson[14], but another alternative is to use the Fisher scoring algorithm. For parameter estimates with an absolute value larger than 0.08, the numerical iteration procedure stops when the relative change in the regression parameter estimates is less than 0.0001 for two successive iterations. For parameter estimates with an absolute value smaller or equal to 0.08, convergence is based on absolute change. The default number of maximum iterations is 50.

SPSS

The GENLIN command in SPSS performs GEE. It can also be selected from the menu using Analyze / Generalized Linear Models / Generalized Estimating Equations. SPSS has implemented the binary coding $Y_{itc} = I(O_{it} = c)$. It also has the option to change the reference category from the highest level to the lowest. Moreover, it provides five options for the working correlation matrix: independent, exchangeable, AR(1), M-independent, and unstructured. SPSS uses Pearson's correlation as the association parameter and applies some functions of Pearson's residual, the so-called Pearson-like residuals, to estimate the association parameters. All of these options can be used with both logit and probit link function.

The robust estimator is the default setting, but a model-based estimator can also be selected. In addition, users are able to choose type I and type III tests using the Wald statistic (default) and the generalized score statistic.

One of the three numerical methods (Newton-Raphson, Fisher scoring and hybrid) for iteration can be chosen. The hybrid procedure is the default. The numerical procedure stops when the absolute change in parameter estimates is less than 0.000001. There is an option to choose for a relative change. The default value for the maximum number of iterations is set to 100, but this can be changed as well.

2.4 Simulation study

2.4.1 Simulation method: multivariate logistic distributions

In our simulation, we used multivariate logistic distributions to generate repeated continuous data first and then changed them to ordinal variables using suitable intervals. Both the k-dimensional logistic Gumbel distribution and a generalization of the k-dimensional Farlie-Gumbel-Morgenstern (FGM) distribution[16] were applied. The standardized multivariate Gumbel distribution has only one parameter θ, and its joint distribution function is given by

$$F_G\left(y_1, ..., y_k\right) = \exp\left[-\left\{ \sum_{i=1}^{k} \left(\log(1 + e^{-y_i}) \right)^{\theta} \right\}^{\frac{1}{\theta}} \right], \tag{2.5}$$

where θ must satisfy: $\theta \geq 1$. This distribution specifies an exchangeable correlation structure among the continuous logistic distributed variables. This means that the correlation of any pair of the k-dimensional vector $(Y_1, ..., Y_k)$ would be the same. Kendall's tau correlation coefficient is given by $\tau = (\theta - 1)/\theta$, but there is no closed form for Pearson's correlation coefficient. Pearson's correlation coefficient can be determined numerically for different values of θ, see Table 2.2. The

Table 2.2 Pearson's correlation coefficient for Gumbel copula.

θ	2	3	4	5
ρ	0.7	0.85	0.91	0.95

standardized (generalization of the) FGM distribution with $2^s - s - 1$ parameters is given by

$$F_{FGM}(y_1, ..., y_k) = \left\{ \prod_{i=1}^{k} \frac{e^{y_i}}{1 + e^{y_i}} \right\} g(y_1, y_2, ..., y_k), \tag{2.6}$$

with

$$g(y_1, y_2, ..., y_k) = 1 + \sum_{j=2}^{k} \sum_{1 \leq r_1 < ... < r_j \leq k} \lambda_{r_1 r_2 ... r_j} \left(\frac{1}{1 + e^{y_{r_1}}}\right)\left(\frac{1}{1 + e^{y_{r_2}}}\right)...\left(\frac{1}{1 + e^{y_{r_j}}}\right),$$

and each parameter must at least satisfy $\left|\lambda_{r_1 r_2 ... r_j}\right| \leq 1$. There are some other restrictions that require that the absolute value of sums of the parameters are also less than one (see Armstrong and Galli,[3]). Choosing different values for elements of the λ's would create an unstructured correlation matrix, i.e. each pairwise correlation of the k-dimensional vector $(Y_1, ..., Y_k)$ could be another value. For example, if the higher order parameters are taken equal to zero and only the bivariate parameters λ_{ij} are non-zero, the correlation of Y_i and Y_j is equal to $3\lambda_{ij}/\pi^2$ $(i \neq j)$.

To be able to simulate from these joint distributions, we applied copula functions. A copula function C links the univariate marginal distributions to their full multivariate distribution[30], i.e.

$$\mathsf{C}\left[F_1(y_1), F_2(y_2), ..., F_k(y_k)\right] = F(y_1, y_2, ..., y_k).$$

If F_L is the standardized univariate logistic distribution, i.e. $F_L(y) = \big(1 +$

$\exp(-y))^{-1}$, then Gumbel copula $\mathsf{C}_G(u_1, ..., u_k)$ is obtained by substituting $F_L^{-1}(u_i)$ for y_i in (2.5). The FGM copula $\mathsf{C}_{FGM}(u_1, ..., u_k)$ is obtained similarly, by substituting $F_L^{-1}(u_i)$ for y_i in (2.6). These copulas are programmed in package copula of R and they can be used to simulate multivariate data (see Yan[48] and Yan et al.[50]).

For the standardized k-dimensional Gumbel logistic distribution, we need to use function archmCopula first and then apply the standard univariate logistic distribution as the marginal distributions for each variable in the function mvdc. The function rMvdc would then generate k-dimensional random variables. For generating a set of k-dimensional variables from the Farlie-Gumbel-Morgenstern distribution, we need to use function fgmCopula first and again apply the standardized univariate logistic distributions as marginal distributions. Generating random variables with function rMvdc for the FGM copula returns a warning which means that the random generation needs to be properly tested. We investigated the generated data and they had the appropriate means and covariances.

For the Gumbel distribution it is easy to generate highly correlated data, since this is determined by the parameter θ directly, see Table 2.2. For the FGM, the restrictions on the parameters imply that the correlations are low. To be able to increase the correlations we added a random intercept variable using the so-called bridge distribution function[45]. The density of this one-parameter distribution function is given by

$$f_b(x) = \frac{1}{2\pi} \frac{\sin(\phi\pi)}{\cosh(\phi x) + \cos(\phi\pi)}, \tag{2.7}$$

with $0 < \phi < 1$ and $-\infty < x < \infty$ and it's distribution function is given by

$$\begin{aligned} F_b(x) &= \frac{1}{\phi\pi \sin(\phi\pi)} \left\{ \arctan\left[\frac{1-\cos(\phi\pi)}{\sin(\phi\pi)}\right] - \right. \\ &\qquad \left. \arctan\left[\frac{\cos(\phi\pi)-1}{\sin(\phi\pi)} \tanh\left(\frac{\phi\pi}{2}\right)\right] \right\}. \end{aligned} \tag{2.8}$$

The mean of the bridge distribution is zero and the variance is $\pi^2\ (\phi^{-2} - 1)/3$. The marginal distribution of the Y's remain of the logistic form in (2.1) and therefore the regression parameters still have an odds ratio interpretation. The bridge distribution was developed for this purpose[45].

The standardized latent variables generated with the multivariate logistic dis-

tributions are not yet related to any covariates. We have selected two explanatory variables time (x_1) and group (x_2) to shift the mean value of the latent variable Y_{it} from zero to η_{it}. We have chosen for η_{it} the following

$$\eta_{it} = -[\beta_T\, x_{1it} + \beta_G\, x_{2i} + \beta_{TG}\, x_{1it} x_{2i}], \tag{2.9}$$

with $\beta_T = 0.5$, $\beta_G = -0.5$, and $\beta_{TG} = -0.5$, with x_{1it} representing time points at which subject i has been observed and with $x_{2i} \in \{0, 1\}$ a group variable. The group variable may represent for instance treatment or gender. Then the shifted latent variable is $Z_{it} = \eta_{it} + Y_{it}$ and the ordinal outcome $O_{it} \in \{1, 2, 3, 4\}$ is created by the cutpoints $\beta_{01} = -1$, $\beta_{02} = 0$, and $\beta_{03} = 1$ to indicate in which of the four intervals $(-\infty, -1]$; $(-1, 0]$; $(0, 1]$; and $(1, \infty)$ the variable Z_{it} is contained.

2.4.2 Simulation method: working correlation matrices

Now consider the binary variable $Y_{itc} = I(O_{it} \leq c) = I(Z_{it} \leq \beta_{0c})$ to transform an ordinal variable to a set of binary variables using the coding of[38]. Using logit link function, Pearson's correlation coefficient between any pair of binary variables Y_{itc} and Y_{itk} within one time point t is given by $\exp(-|\beta_{0c} - \beta_{0k}|/2)$ and does not depend on time nor on any other covariate (see Kenward et al.[15]). This would also be true for the coding $Y_{itc} = I(O_{it} > c)$ but this is not the case for the coding $Y_{itc} = I(O_{it} = c)$.

Indeed, the correlation coefficient between $I(O_{it} = c)$ and $I(O_{it} = k)$ is determined by the multinomial distribution function and it is of the form $-[\mu_{itc}\mu_{itk}/(1-\mu_{itc})(1-\mu_{itk})]^{1/2}$, with μ_{itc}, in terms of the general model (2.1), given by

$$\mu_{itc} = \frac{\exp\left[\beta_{0c} + x_{it}^T\beta\right]}{1 + \exp\left[\beta_{0c} + x_{it}^T\beta\right]} - \frac{\exp\left[\beta_{0c-1} + x_{it}^T\beta\right]}{1 + \exp\left[\beta_{0c-1} + x_{it}^T\beta\right]}.$$

Hence, the correlation between the coded binary variables $I(O_{it} = c)$ and $I(O_{it} = k)$ do depend on time t and on subject specific values of the covariates for subject i. Thus as soon as the simulated variables were affected by covariates, the Pearson correlation between the coded binary variables $Y_{itc} = I(O_{it} = c)$ would depend on the covariate time t and subject specific variables. Despite the difference in correlation coefficients between these choices of coding, sofar it all fits nicely with the theory and the software packages we are studying. The software packages have implemented these exact same correlations.

If we now investigate the correlation of the binary variables (Y_{itc}, Y_{isk}) from different time points t and s, with $t \neq s$, Pearson's correlation coefficient is given by

$$\mathsf{CORR}(Y_{itc}, Y_{isk}) = \frac{F_{ts}\left(\beta_{0c}, \beta_{0k}\right) - F_t\left(\beta_{0c}\right) F_s\left(\beta_{0k}\right)}{\left\{F_t\left(\beta_{0c}\right)\left[1 - F_t\left(\beta_{0c}\right)\right] F_s\left(\beta_{0k}\right)\left[1 - F_s\left(\beta_{0k}\right)\right]\right\}^{\frac{1}{2}}}, \tag{2.10}$$

with $F_{ts}\left(\beta_{0c}, \beta_{0k}\right)$ the bivariate distribution of the latent variables Z_{it} and Z_{is} evaluated in (β_{0c}, β_{0k}). Note that the correlation in (2.10) holds when $c = k$ and $c \neq k$. This correlation coefficient will depend on the time points t and s through the means η_{it} and η_{is} of Z_{it} and Z_{is} in (2.9) and through the correlation of the latent variables Z_{it} and Z_{is}. Thus for the Gumbel distribution, the correlations of the latent variables Z_{it} and Z_{is} are exchangeable but the correlation between the binary variables Y_{itc} and Y_{isk} are not exchangeable due to the time dependent mean in (2.9). Furthermore, the correlation in (2.10) does also depend on the subject specific variables of the covariates used in (2.9). This means that the correlation of the binary variables Y_{itc} and Y_{isk} would still depend on the subjects whenever the mean (2.9) would include covariates even if all subjects would have been observed at the exact same time points. This implies that the Gumbel and the FGM distribution using the mean in (2.9) would never impose the correct weight matrix or working correlation matrix for multgee, geepack, repolr, SAS and SPSS, due to the use of covariates in the latent variables. Note that this would remain true when we would use $I(O_{it} = c)$ or $I(O_{it} > c)$ for binary variables. Therefore, the correlation between binary variables can be quite different from the correlation between the latent variables[29].

Only in case the mean in (2.9) would be independent of the covariates and time, i.e. $\eta_{it} = \eta$ is constant, then the Gumbel distribution would provide the exact exchangeable working correlation used in SPSS. Indeed, the correlation coefficient in (2.10) for $I(O_{it} = c)$ is only a function of the cutpoints β_{0c} and β_{0k}, say ρ_{ck}, and independent of time. However, the correlation coefficient in (2.10) for the Gumbel distribution would not generate the uniform working correlation matrix of repolr, since they have implemented the form $\rho_{ck} = \alpha \exp(-|\beta_{0c} - \beta_{0k}|/2)$ which is unequal to (2.10). When the mean would be constant, any exchangeable multivariate latent variable would lead to the exchangeable working correlation matrix used in SPSS, but only a specific subset of this class of multivariate distributions would generate the uniform working correlation matrix of repolr. The mis-specification of the

correlation matrices when (2.9) is used should not be an issue though, since GEE should still lead to the correct estimates of the parameters β when the mean in (2.1) is correctly specified[28].

2.4.3 Simulation method: parameter settings

We simulated three dimensional logistic distributed variables (three time points) using parameters $\theta \in \{2, 3, 4, 5\}$ for the Gumbel distribution (G_θ) and parameters $\lambda_{12} = 0.3$, $\lambda_{13} = -0.3$, $\lambda_{23} = 0.3$, and $\lambda_{123} = 0$ for the FGM distribution in combination with a bridge distribution having parameter $\phi = \sqrt{0.5}$. Pearson's correlation coefficients of the three-dimensional latent variable will be in the interval $(0.45, 0.55)$ for this selected FGM distribution with bridge distribution. We used the same number of subjects for both levels of the group variable x_2. Furthermore, we used constant time points for all subjects, $x_{1i1} = 0$, $x_{1i2} = 1$, and $x_{1i3} = 2$ and time varying time points $x_{1i1} = 0$, $x_{1i2} = 1 + u_{i2}$, and $x_{1i3} = 2 + u_{i3}$, with u_{i2} and u_{i3} independently generated using the uniform distribution on $(-0.25, 0.25)$. The time varying time points were included to simulate data that is not perfectly balanced for the time and group variable. For binary data it has been demonstrated that the independent working correlation matrix is as efficient as the exchangeable working correlation matrix in such balanced settings[25].

We also investigated the effect of some variance heterogeneity in the latent variable. The heterogeneity is determined by multiplying the latent variable, before shifting it with mean (2.9), at each time point with $\exp\{-0.1\, x_{2i}\}$. This means that the group of subjects with $x_{2i} = 1$ has a decreased variance of approximately 20% at each time point compared to the subjects with $x_{2i} = 0$. This variance heterogeneity does not affect the mean in (2.9) for the shifted latent variable Z_{it}, since only the variance of the latent variable has changed, but it does effect the linear relationship in (2.1) for the ordinal outcome variable.

The relationship becomes of the following form $\text{logit}[\mathbb{P}(O_{it} \leq c)] = \beta_{0c} + \beta_T x_{1it} + \beta_{2c} x_{2it} + \beta_{12} x_{1it} x_{2it}$, with $\beta_{2c} = (\exp\{0.1\} - 1)\beta_{0c} + \beta_G \exp\{0.1\}$ and $\beta_{12} = (\exp\{0.1\} - 1)\beta_T + \beta_{TG} \exp\{0.1\}$, and β_T, β_G, and β_{TG} defined in (2.9). It destroys the assumption of proportionality for the main effect of the variable x_{2i}. Thus in this setting the linear relationship in (2.1) is incorrectly specified, although we have correctly selected the right variables and the linear relationship in (2.9) is still true for the latent variables. Thus both the mean in the logit

scale and the working association matrix would be incorrectly specified when this variance heterogeneity is present in the latent variable. We evaluated how well the proportionality test, which is available in repolr only, would pick up this small violation compared to the other models for which proportionality would be true.

An overview of the parameter settings we used to simulate data are provided in the Table 2.3. The Gumbel distribution with $\theta = 3$ includes the heterogeneity in variance for the group variable. For each parameter setting, 1000 simulated data sets were determined. Each data set was analysed with R, SAS, and SPSS. On each data set, we applied "independence", "exchangeable", and "unstructured" association structures with each package (if available).

Table 2.3 Overview of settings for the simulations study.

Number of subjects	Time points are the same	Time points vary with individuals
15 subjects per groups	G_2; FGM	-
50 subjects per groups	G_5; FGM	G_2; G_3; G_4; G_5; FGM
150 subjects per groups	G_4	-

2.5 Results

2.5.1 Theoretical considerations

As mentioned in Section 2.3.1, SAS and SPSS work with the same set of coded binary variables that is also chosen in the approach of Lipsitz et al.[21]. They also used Pearson's correlation coefficients to estimate of the working correlation. The difference in signs in the parameters estimates, shown in the example, is due to SPSS, which uses $\beta_{0c} - x_{it}^\top \beta_1$ in (2.1) rather than $\beta_{0c} + x_{it}^\top \beta_1$, which is used by other packages.

Recall that there were three ways of coding the ordinal outcomes into binary variables[21,44,11,38] and that there were different estimation approaches for the association parameters. The difference in coding should, in principle, not lead to different estimates of the fixed parameters β, since there a linear transformation exists between the binary variables. If Y_{it}^L is a vector of binary variables with $Y_{itc}^L = I(O_{it} = c)$ and Y_{it}^P is a vector with $Y_{itc}^P = I(O_{it} \leq c)$, then $Y_{it}^L = GY_{it}^P$, with G a $(C-1) \times (C-1)$ lower triangular matrix. The matrix G is invertible since all the elements on the main diagonal are equal to one. This unique relationship between the binary variables connects the means and covariances between the

two sets of binary variables: i.e. $\mu_i^L = G\mu_i^P$ and $V_i^L = GV_i^P G^\top$. Hence, under these two sets of coding the GEE equations solve the same set of parameters β. This would hold true with any two pair-wise comparisons of the three different theories. However, different parameterisations of the association parameters and different methods for estimation of these parameters could still lead to different estimation of α, and thus of β.

The parameter estimates generated by R-geepack in our case study compared to the estimates of the other packages (under independence) is awkward considering the linear transformation in binary variables. We could not identify what causes the difference, but we did observe an inconsistency with Pearson's correlation coefficient for the association of binary variables within one time point. The correlation coefficient between any pair $Y_{itc} = I(O_{it} > c)$ and $Y_{itk} = I(O_{it} > k)$ with the choice of the logit link function, results into $\exp(-|\beta_{0c} - \beta_{0k}|/2)$. For a small and simple data set, where the ordinal outcome has only three levels and two time points and were generated independently, we substituted the parameter estimates of ordgee in the GEE's. We used the incorrect value $\exp(|\beta_{0c} - \beta_{0k}|/2)$ for the correlation coefficient of Y_{itc} and Y_{itk} with the independence association structure and we obtained a set of equations that is almost equal to zero. This did not occur when we used the correct estimated correlation coefficients.

2.5.2 Simulation results

Due to the unexpected results of function ordgee in the geepack and the above explanation, we decided to exclude this package from our simulation study. The results of SAS were also excluded since it has only one option for the working correlation matrix. Furthermore, the SAS and SPSS results are the same with this choice of working correlation matrix (except of course for the difference in regression parameter signs). Hence, this simulation study provides the results of SPSS, repolr and multgee using the logit link function under independence, time exchangeability and unstructured association structures.

Simulation results: numerical convergence

Convergence problems were observed with SPSS, repolr, and multgee. Tables 2.4 and 2.5 present an overview of the percentages of simulation runs with numerical issues for each of the simulation studies.

Table 2.4 Percentages of convergence issues for simulated data sets with constant time points for individuals.

Analysis approach		G_2	G_3	G_4	G_5	FGM	FGM
		n = 30	n = 100	n = 300	n = 100	n = 30	n = 100
Exchangeable	SPSS	3.4%	0.01%	0	0.01%	1.5%	0
	repolr	14.1%	6.6%	0	4.3%	1.9%	0
	multgee	0.4%	0	0	0	0	0
Unstructured	SPSS	26.7%	3.7%	0.01	11.2%	17.8%	0
	multgee	32.2%	1.6%	0	16.2%	10.2%	0

For constant time points, repolr has more numerical issues than multgee and SPSS when the exchangeable working association matrix is selected. It seems that the numerical issues for replor are related to the number of subjects. For the choice of coded binary variables used in repolr, the correlation coefficient across time gets closer to the exchangeable structure and is more alike for each subject as the parameter θ of the Gumbel distribution decreases. This means that we expect fewer numerical issues at $\theta = 2$ than at $\theta = 5$. The relatively high percentage of 14.1% at $\theta = 2$ compared to the percentage 4.3% at $\theta = 5$, suggests that the difference in sample size may play a role in the convergence. For the coding of binary variables used in SPSS all selected Gumbel distributions will yield working correlation matrices for the association over time that are quite different from exchangeability and are not alike for subjects from the different groups. Numerical issues with SPSS are limited to the choice of unstructured working correlation matrix. When the total number subjects is relatively small, substantial numbers of data sets will demonstrate numerical issues. Note that the unstructured working correlation matrix in SPSS requires 27 correlation coefficients alone to describe all of the association parameters across time, which is substantial for this relative small set of subjects. Multgee shows the lowest numbers of non-convergence with respect to repolr and SPSS for exchangeability, but it performs worse than SPSS for G_2 and G_5 when the unstructured association is applied. Similar to SPSS, sample size appears to affect the convergence for an unstructured association. This may be caused by estimated local odds ratios close to the boundary values[44].

For varying time points we do see that the numerical issues of repolr and multgee increase with the parameter θ of the Gumbel distribution. Note that the number of subjects is constant at 100. This is probably explained by the fact that the working association diverges from exchangeability when θ increases. Compar-

Table 2.5 Percentages of convergence issues for simulated data sets with time varying time points for individuals.

Analysis approach		G_2	G_3	G_4	G_5	FGM
		n = 100	n = 100	n = 100	n = 100	n = 100
Exchangeable	SPSS	28.2%	5.6%	2.5%	0.8%	0
	repolr	0.8%	3.6%	9.0%	5.4%	0
	multgee	0	0	0.01%	0.2%	0
Unstructured	SPSS	22.8%	9.24%	10.7%	14.2%	0.9%
	multgee	0	1.2%	8.2 %	20.9%	0

ing Gumbel distribution G_5 from Table 2.5 with the same distribution in Table 2.4, suggests that the varying time points for subjects enhance the numerical issues for repolr and multgee, because the working association matrix is now different for each subject due to the subject specific time points. For varying time points SPSS has substantial numerical issues when the unstructured working correlation matrix is selected, but also with the exchangeable working correlation matrix for Gumbel G_2. We suspect that the true correlation matrix for the ordinal outcomes, that is induced by the latent variables strongly, mismatch with the choice of working correlation matrix that is estimated. This may indicate that the estimated working correlation matrix is not a genuine correlation matrix, which may imply that the estimated correlation matrix is not invertible (see Chaganty and Joe[5,6]). This applies to for both SPSS and repolr, because they both reported a problem with the covariance matrix V_i for most of their numerical issues (repolr: "grad2 $<$ 0 minimum for alpha not achieved"; SPSS: "the Hessian matrix is singular"). Unfortunately, multgee does not provide any clear information useful to diagnose the non-convergence issue. Changing some of the default settings, such as relaxing the criteria of the convergence of the parameter estimates or the maximum number of iterations, solved only a few of the numerical issues.

2.5.3 Simulation results: parameter estimates

When the working association matrix is selected as independence or exchangeable and the number of subjects is moderate to large (say 100 subjects or more), most of the simulation cases provided a bias in the parameter estimates up to a few percent (0−4%). This is illustrated for the Gumbel distribution with $\theta = 5$ for 100 subjects and $\theta = 4$ with 300 subjects given in Tables 2.6 and 2.7, respectively. Furthermore, recall that the true parameters for G_5 are equal to $\beta_{01} = -1$, $\beta_{02} = 0$, $\beta_{03} = 1$,

$\beta_T = 0.5$, and $\beta_G = \beta_{TG} = -0.5$. Significant bias is obtained for the simulations

Table 2.6 The mean parameter estimates (standard error) of 1000 simulated data sets, from the Gumbel distribution with $\theta = 5$, 100 subjects, and constant time points.

	Independence	Exchangeable			Unstructured	
	All*	SPSS	repolr	multgee	SPSS	multgee
β_{01}	-1.04(0.28)	-1.04(0.32)	-1.04(0.28)	-1.03(0.28)	-0.97(0.30)	-1.04(0.28)
β_{02}	-0.01(0.26)	0.00(0.32)	-0.01(0.26)	0.00(0.26)	0.04(0.29)	-0.01(0.26)
β_{03}	1.01(0.28)	1.02(0.34)	1.01(0.28)	1.02(0.28)	1.06(0.32)	1.01(0.28)
β_T	0.51(0.08)	-0.50(0.10)	0.51(0.08)	0.51(0.08)	-0.49(0.09)	0.50(0.08)
β_G	-0.50(0.36)	0.51(0.41)	-0.49(0.36)	-0.51(0.36)	0.59(0.40)	-0.52(0.36)
β_{TG}	-0.51(0.10)	0.51(0.12)	-0.51(0.10)	-0.50(0.10)	0.48(0.13)	-0.48(0.10)

* SPSS gives the same values but with opposite signs on β_T , β_G, and β_{TG}.

Table 2.7 The mean parameter estimates (standard error) of 1000 simulated data sets, from the Gumbel distribution with $\theta = 4$, 300 subjects, and constant time points.

	Independence	Exchangeable			Unstructured	
	All*	SPSS	repolr	multgee	SPSS	multgee
β_{01}	-1.01(0.15)	-1.01(0.18)	-1.01(0.16)	-1.01(0.16)	-1.00(0.17)	-1.01(0.16)
β_{02}	-0.01(0.15)	0.00(0.19)	-0.01(0.15)	-0.01(0.15)	0.01(0.17)	0.00(0.15)
β_{03}	1.00(0.16)	1.01(0.19)	1.00(0.16)	1.00(0.16)	1.01(0.18)	1.00(0.16)
β_T	0.50(0.05)	-0.50(0.06)	0.50(0.05)	0.50(0.05)	-0.50(0.06)	0.50(0.05)
β_G	-0.49(0.21)	0.50(0.23)	-0.49(0.21)	-0.50(0.21)	0.52(0.22)	-0.50(0.21)
β_{TG}	-0.51(0.06)	0.50(0.07)	-0.51(0.06)	-0.50(0.06)	0.50(0.07)	-0.49(0.06)

* SPSS gives the same values but with opposite signs on β_T , β_G, and β_{TG}.

with only 30 subjects using independence or exchangeability. The bias increases to approximately 10% for the Gumbel distribution with $\theta = 2$ (Table 2.8) and to 17.2% for the FGM with constant time points across individuals (Table 2.9). Note that the true parameter values for the FGM distribution, combined with the bridge distribution, are equal to $\beta_{01} = -0.71$, $\beta_{02} = 0$, $\beta_{03} = 0.71$, $\beta_T = 0.35$, and $\beta_G = \beta_{TG} = -0.35$. This change in true value compared to the Gumbel distribution is caused by the additional variation of the random intercept value that follows the bridge distribution. When the unstructured working association is used, even higher biases are obtained. The bias with multgee can be as high as 20% for the FGM distribution combined with the bridge distribution (Table 2.9) while SPSS yields biases of 40% for both G_2 distribution (Table 2.8) and FGM distribution (Table 2.9). When the number of subjects increases to 100, the biases with multgee and SPSS under the unstructured association reduce to values as high as 4% and 18%, respectively (Table 2.6). When the number of subjects increases further to 300, the biases with multgee and SPSS almost vanish(Table

2.7).

Table 2.8 The mean parameter estimates (standard error) of 1000 simulated data sets, for the Gumbel distribution with $\theta = 2$, 30 subjects and constant time point.

	Independence	Exchangeable			Unstructured	
	All*	SPSS	repolr	multgee	SPSS	multgee
β_{01}	-1.05(0.48)	-1.05(0.55)	-1.06(0.48)	-1.04(0.47)	-0.91(0.52)	-1.03(0.47)
β_{02}	0.00(0.46)	0.01(0.55)	-0.01(0.47)	0.00(0.46)	0.12(0.52)	0.00(0.45)
β_{03}	1.04(0.49)	1.07(0.58)	1.04(0.49)	1.05(0.49)	1.20(0.60)	1.05(0.49)
β_T	0.53(0.22)	-0.51(0.26)	0.51(0.22)	0.51(0.21)	-0.45(0.28)	0.48(0.21)
β_G	-0.52(0.65)	0.55(0.71)	-0.51(0.65)	-0.53(0.64)	0.70(0.76)	-0.55(0.63)
β_{TG}	-0.54(0.30)	0.53(0.34)	-0.53(0.31)	-0.51(0.29)	0.43(0.42)	-0.47(0.29)

* SPSS gives the same values but with opposite signs on β_T , β_G, and β_{TG}.

Table 2.9 The mean parameter estimates (standard error) of 1000 simulated data sets, for the FGM with the bridge distribution for 30 subjects and constant time point.

	Independence	Exchangeable			Unstructured	
	All*	SPSS	repolr	multgee	SPSS	multgee
β_{01}	-0.73(0.47)	-0.73(0.48)	-0.73(0.47)	-0.74(0.47)	-0.67(0.53)	-0.74(0.46)
β_{02}	0.01(0.46)	0.01(0.47)	0.01(0.46)	0.00(0.46)	0.08(0.52)	0.01(0.45)
β_{03}	0.76(0.47)	0.77(0.49)	0.77(0.47)	0.75(0.47)	0.86(0.59)	0.76(0.46)
β_T	0.36(0.27)	-0.36(0.28)	0.36(0.27)	0.36(0.26)	-0.32(0.32)	0.34(0.25)
β_G	-0.40(0.64)	0.41(0.65)	-0.41(0.64)	-0.40(0.64)	0.49(0.74)	-0.42(0.62)
β_{TG}	-0.35(0.38)	0.35(0.39)	-0.35(0.38)	-0.34(0.38)	0.29(0.47)	-0.30(0.36)

* SPSS gives the same values but with opposite signs on β_T , β_G, and β_{TG}.

The bias results presented here were obtained from using simulations with constant time points across subjects. With varying time points, the bias was similar to the results for constant time points, except for SPSS, which gave biases as high as 16% for the parameter estimate of the group variable with the exchangeable working correlation matrix. For unstructured association this increased to 26%. These inflation in biases were not seen with repolr and multgee.

Bias could possibly affect the coverage probability of the Wald confidence interval on the parameter. Absence of bias however, does not guarantee a nominal level of coverage, since the standard error of the parameter estimate would also affect the coverage probability. Tables 2.10 and 2.11 provide coverage probabilities for two simulation settings.

Overall, the coverage probabilities were relatively close to the nominal value of 95% when either the independence or exchangeable working association matrix was used. The coverage probabilities for the independence working association matrix ranged from 93.0% to 96.5%, and for exchangeability it ranged from 92.0%

Table 2.10 The percentage of coverage probability for the FGM distribution combined with bridge distribution with 30 subjects and constant time points.

	Independence			Exchangeable			Unstructured	
	SPSS	repolr	multgee	SPSS	repolr	multgee	SPSS	multgee
β_{01}	95.4	95.4	95.4	95.4	95.5	95.1	92.8	95.1
β_{02}	95.0	95.0	95.0	95.3	95.0	95.3	92.7	95.8
β_{03}	94.7	94.7	94.7	95.0	94.5	94.9	92.3	93.8
β_T	93.2	93.0	93.2	93.0	93.1	92.6	90.0	90.4
β_G	93.9	93.9	93.9	93.4	94.2	93.6	91.1	93.9
β_{TG}	95.1	95.1	95.1	95.5	95.0	94.9	90.4	93.4

Table 2.11 The percentage of coverage probability for the Gumbel distribution with $\theta = 5$ and 100 subject with time varying.

	Independence			Exchangeable			Unstructured	
	SPSS	repolr	multgee	SPSS	repolr	multgee	SPSS	multgee
β_{01}	95.1	94.7	95.1	94.8	94.8	94.2	91.0	93.8
β_{02}	94.8	94.3	94.8	93.4	95.0	94.2	92.2	94.1
β_{03}	94.3	93.9	94.3	95.0	94.4	93.9	93.1	93.5
β_T	94.1	93.5	94.1	92.5	93.7	92.7	89.9	92.2
β_G	95.6	95.6	95.6	95.9	95.6	95.1	94.1	94.8
β_{TG}	94.2	94.9	94.2	95.6	94.9	94.2	92.3	92.5

to 96.4%. The coverage probability has a tendency to be somewhat more liberal than being conservative for smaller sample sizes. In general however, the standard errors seem to be estimated at the correct level, despite the difference between the selected working association matrices and the true correlation matrix induced by the latent variable models, and the coverage probabilities are reasonably close to the nominal value.

For the unstructured association matrix we obtained in some cases poor coverage probabilities (see Tables 2.10 and 2.11). The observed range of coverage probabilities for the unstructured working association matrix was equal to 89.7% to 95.8% for all our simulation settings, which demonstrates liberal coverage probabilities. It should be noted that the bias for the parameter β_T in Table 2.10 was only 2.8% for multgee, but the coverage probability was still only 90.4%. This may indicate an underestimated standard error for this parameter when the unstructured association matrix is used. This issue has been reported before by[17], who indicated that these issues may be related to small sample sizes. On the other hand, this setting also demonstrated substantial numbers of simulated data sets with non-convergence, which could also be an explanation of the poor coverage.

Table 2.12 The percentage of coverage probability for the Gumbel distribution with $\theta = 4$ and 300 subject with time constant.

	Independence			Exchangeable			Unstructured	
	SPSS	repolr	multgee	SPSS	repolr	multgee	SPSS	multgee
β_{01}	94.3	94.3	94.3	94.3	94.2	94.7	93.8	94.5
β_{02}	94.1	94.0	94.0	94.7	94.1	94.1	94.3	94.1
β_{03}	94.3	94.3	94.3	94.1	94.2	93.6	93.9	93.7
β_T	93.6	93.6	93.6	94.9	93.8	93.9	94.3	94.3
β_G	95.2	95.2	95.2	95.1	95.2	95.2	95.2	95.4
β_{TG}	95.0	95.0	95.0	95.0	94.9	94.6	94.2	93.5

2.5.4 Simulation results: test for proportionality

Only R-repolr has a test for the assumption of proportionality and we investigated the test in our simulation. The proportionality assumption was guaranteed for all our latent variable models, except for the Gumbel distribution with $\theta = 3$, when we used constant time points for the 100 subjects. To violate the proportionality assumption, we introduced variance heterogeneity for the group variable. The rejection rates are presented in Table 2.13.

Table 2.13 Percentages of rejection the proportionality assumption in the simulated data sets.

	G_2	G_2	G_3	G_4	G_4	G_5	FGM	FGM
	n=30	n=100	n=100	n=100	n=300	n=100	n=30	n=100
Time points fixed	14.3%	NA	11.4%	NA	11.1%	16.1%	15.5%	5.2%
Time points varying	NA	8.5%	11.6%	14.3%	NA	13.7%	NA	5.9%

The proportionality test has an inflated type I rejection rate, since it frequently rejected the proportionality assumption more than 5%, which was the selected level of significance. The type I error rate was close to 5%, only for the FGM distribution combined with the bridge distribution using 100 subjects. Only 11.4% of the simulated data sets with a violated proportionality assumption was detected using heterogeneity setting. This is as large as the false rejection rates for situations where the proportionality is true. The difference in variance in the latent variable between the two groups of subjects ($x_2 = 0$ versus $x_2 = 1$) was only 20%. This could have been too small to be detected with the proportionality test, although a 20% change in variance could be clinically relevant.

2.6 Conclusions and recommendations

In this paper, we reviewed the GEE method for longitudinal ordinal data with five software packages: SAS 9.4 (GENMOD procedure), SPSS 22.0.0 (GENLIN command), repolr (function repolr), multgee (function ordLORgee), and geepack (function ordgee) in R 3.0.2. We saw that SPSS implements a minus sign for the regression parameters, implying a different interpretation of the coefficients. If an explanatory variable has a positive effect on the response, SPSS estimates a negative effect and vice versa. SAS provides only the independence working correlation matrix available, while SPSS is flexible in offering other types of working correlation matrices. Within R, the geepack, multgee and repolr packages all use a different set of binary variables for coding ordinal data. SAS, SPSS, and multgee use the same coding. We demonstrated theoretically that this should not lead to different parameter estimates, since the coding of the ordinal outcome into binary variables are linearly related. The numerical procedures and the method of estimation of the association parameters could lead to possible differences, however, particularly for the estimation of the standard errors. Ultimately, geepack provided results that differ significantly from the other packages. The parameter estimates may give rise to a correlation coefficient for the coded binary variables within a time point that is larger than one. The final comparison included SPSS, multgee, and repolr.

We applied copula functions for generating multivariate logistic distributed latent variables underneath the ordinal outcomes. This provided us with a completely specified multivariate distribution on which we could evaluate the performance of the software packages. The covariates in the mean of the latent variables induce a correlation matrix for the ordinal outcomes across time that does not seem to fit with the current choices of the software packages. The reason is that the covariates change the associations across subjects for time dependent association, while SPSS, repolr, and multgee fit constant associations for subjects. Moreover, correlation of coded binary outcomes across time must satisfy a specific exponential form for repolr which is not required for multgee and SPSS when both packages use the exchangeable associations. Mis-specification of the working correlation should however not be any problem since GEE should still be able to estimate the mean parameters appropriately when the robust estimator is applied.

The sample size for subjects has a strong effect on the numerical process.

SPSS, repolr, and multgee show non-convergence issues when we apply the non-independence association matrices. When we used time points that were exactly the same for each subject and the exchangeable working association matrix, SPSS appeared to have fewer numerical issues than repolr and multgee seems to have less numerical issues than SPSS. Multgee with exchangeable association is again the best for varying time points but no clear conclusion could be given between SPSS and repolr. For some settings repolr was (substantially) better, but in other settings SPSS seem to do better again. The unstructured association in SPSS and multgee, which is not offered by repolr, seem to cause the largest problems with numerical convergence. Multgee seems to do better than SPSS, although multgee was worse than SPSS for settings with large correlations among the latent variables. Most numerical issues in SPSS were related to problems with non-invertible covariance matrices. It could almost never be resolved by changing the default settings. On the other hand, multgee did not provide clear information on the numerical issues at all.

For relative small numbers of subjects, say 30 subjects in total, we saw the largest bias in the parameter estimates. This bias declined to less than 4% when the number of subjects increased to more than 100 and the association matrix was equal to independence or exchangeable. The unstructured association matrix needed even more subjects ($N = 300$) to reduce the bias to acceptable levels on all parameters for both packages (although multgee had less bias than SPSS). The coverage probabilities on the mean parameters were relatively close to the nominal value for both the independence and exchangeable association matrices, although they would be more liberal than conservative. The coverage probabilities ranged from 93.0% to 96.5% and from 92.0% to 96.4%, respectively. The unstructured association matrix showed mostly liberal coverage probabilities, even for larger sample sizes and for settings without strong biases. The coverage probabilities ranged from 89.7% to 95.8%.

R-repolr, which can only fit the cumulative logistic link function, has an advantage over SPSS and multgee because it contains an option for testing the proportionality assumption. Unfortunately, the type I error rate of this test is highly inflated, which is consistent with results published on logistic regression[1]. When we introduced a simulation setting that has a violated proportionality assumption, by introducing variance heterogeneity in the latent variables across subjects, the proportionality test did not find this violation frequently. This has to do with the

size of the variance heterogeneity, which was in our setting at 20% difference, i.e. one group had a 20% lower variance than the other group systematically at across time points. We conclude that this test should be applied with care.

Overall, the GEE method performed quite well when there are more than 100 subjects and the association structure is not too complex (no unstructured). GEE produced limited bias and the numerical procedures work appropriately. When sample size is small or proportionality is violated, the methods may result in biased estimates and more numerical issues. In all our simulated data sets the independence association structure with the robust estimator performed quite well and can be used at least as starting point for the analysis. Both SPSS and repolr are recommended but in our simulation multgee outperformed SPSS and repolr in parameter estimation. Both SPSS and multgee provide more flexibility in their choice of association and allow different link functions, but repolr allows proportionality testing. Additional research on GEE is needed, in particular on addressing the subject-specific correlation of the ordinal outcomes across time, and for improving methods on detecting issues with the assumption of proportionality.

Acknowledgement

We thank Ms. Elizabeth Groom for reading our manuscript and improving the English, but we take full responsibility if any mistakes are still included.

References

[1] Agresti, A. (2010). *Analysis of ordinal categorical data.* John Wiley & Sons, 2^{nd} edition edition.

[2] Agresti, A. and Natarajan, R. (2001). Modeling clustered ordered categorical data: A survey. *International Statistical Review*, 69:345–371.

[3] Armstrong, M. and Galli, A. (2002). Sequential nongaussian simulations using the fgm copula.

[4] Carey, V., Zeger, S., and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, 80(3):517–526.

[5] Chaganty, N. and Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society B*, 66(4):851–860.

[6] Chaganty, N. and Joe, H. (2006). Range of correlation matrices for dependent bernoulli random variables. *Biometrika*, 93(1):197–206.

[7] Clayton, D. (1992). Repeated ordinal measurements: A generalised estimating equation approach.

[8] Deming, W. and Stephan, F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11:427–444.

[9] Goodman, L. (1985). The analysis of cross-classified data having ordered and /or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics*, 13:10–69.

[10] Hardin, J. and Hilbe, J. (2012). *Generalized estimating equations.* Chapman and Hall/CRC Press, 2^{nd} edition edition.

[11] Heagerty, P. and Zeger, S. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, 19(435):1024–1036.

[12] Højsgaard, S., Halekoh, U., and Yan, J. (2005). The *r* package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2):1–11.

[13] Horton, N. and Lipsitz, S. (1999). Review of software to fit generalized estimating equation regression models. *The American Statistician*, 53(2):160.

[14] Jennrich, R. and Sampson, P. (1976). Newton-raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, 18(1):11–17.

[15] Kenward, M., Lesaffre, E., and Molenberghs, G. (1994). An 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(4):945–953.

[16] Kotz, S., Balakrishnan, N., and Johnson, N. (2000). *Continuous multivariate distributions, models and applications*. John Wiley & Sons, 2^{nd} edition edition.

[17] Li, Y. and Schafer, D. (2008). Likelihood analysis of the multivariate ordinal probit regression model for repeated ordinal responses. *Computational Statistics & Data Analysis*, 52(7):3474–3492.

[18] Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

[19] Liang, K., Zeger, S., and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society B*, 54(1):3–40.

[20] Lipsitz, S. and Fitzmaurice, G. (1996). Estimating equations for measures of association between repeated binary responses. *Biometrics*, 52(3):903–912.

[21] Lipsitz, S., Kim, K., and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, 13(11):1149–1163.

[22] Lipsitz, S., Laird, N., and Harrington, D. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika*, 78(1):153–160.

[23] Liu, I. and Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test*, 14(1):1–73.

[24] Lumley, T. (1996). Generalized estimating equations for ordinal data: A note on working correlation structures. *Biometrics*, 52(1):354–361.

[25] Mancl, L. and Leroux, B. (1996). Efficiency of regression estimates for clustered data. *Biometrics*, 52(2):500–511.

[26] McCullacgh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society B*, 42(2):109–142.

[27] Miller, M., Davis, C., and Landis, J. (1993). The analysis of longitudinal polytomous data: generalized estimating equations and connections with weighted least squares. *Biometrics*, 49(4):1033–1044.

[28] Molenberghs, G. and Kenward, M. (2010). Semi-parametric marginal models for hierarchical data and their corresponding full models. *Computional Statistics & Data Analysis*, 54:585–597.

[29] Molenberghs, G. and Verbeke, G. (2005). *Models for discrete longitudinal data*. Springer-Verlag.

[30] Nelson, R. (2006). *An introduction to copula*. Springer-Verlag, 2^{nd} edition edition.

[31] Nores, M. and del Pilar Díaz, M. (2008). Some properties of regression estimators in gee models for clustered ordinal data. *Computational Statistics & Data Analysis*, 52(7):3877–3888.

[32] Oster, R. (2002). An examination of statistical software packages for categorical data analysis using exact methods. *The American Statistician*, 56(3):235–246.

[33] Oster, R. (2003). An examination of statistical software packages for categorical data analysis using exact methods-part ii. *The American Statistician*, 57(3):201–213.

[34] Oster, R. and Hilbe, J. (2008). An examination of statistical software packages for parametric and nonparametric data analyses using exact methods. *The American Statistician*, 62(1):74–84.

[35] Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57(1):120–125.

[36] Parsons, N. (2012). *repolr: Repeated measures proportional odds logistic regression*. R package version 1.0.

[37] Parsons, N., Costa, M., Achten, J., and Stallard, N. (2009). Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package r. *Computational Statistics & Data Analysis*, 53(3):632–641.

[38] Parsons, N., Edmondson, R., and Gilmour, S. (2006). A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research. *Journal of the Royal Statistical Society C*, 55(5):507–524.

[39] Prentice, R. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44(4):1033–1048.

[40] Qu, Y., Williams, G., Beck, G., and Medendorp, S. (1992). Latent variable models for clustered dichotomous data with multiple subclusters. *Biometrics*, 48:1095–1102.

[41] Stiger, T., Barnhart, H. X., and Williamson, J. (1999). Testing proportionality in the proportional odds model fitted with gee. *Statistics in Medicine*, 18(11):1419–1433.

[42] Stokes, M., Davis, C., and Koch, G. (2000). *Categorical data analysis using the SAS system*. SAS Institute, 2^{nd} edition edition.

[43] Touloumis, A. (2013). multgee: Gee solver for correlated nominal or ordinal multinomial responses.

[44] Touloumis, A., Agresti, A., and Kateri, M. (2013). Gee for multinomial responses using a local odds ratios parameterization. *Biometrics*, 69:633–640.

[45] Wang, Z. and Louis, T. (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika*, 90(4):765–775.

[46] Williamson, J., Kim, K., and Lipsitz, S. (1995). Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, 90(432):1432–1437.

[47] Williamson, J., Lipsitz, S., and Kim, K. (1999). Geecat and geegor: Computer programs for the analysis of correlated categorical response data. *Computer Methods and Programs in Biomedicine*, 58:25–34.

[48] Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):1–21.

[49] Yan, J., Højsgaard, S., and Halekoh, U. (2012a). geepack: Generalized estimating equation package. R package version 1.1-6.

[50] Yan, J., Kojadinovic, I., Hofert, M., and Maechler, M. (2012b). *copula: Multivariate dependence with copulas.* R package version 0.99-1.

[51] Yu, K. and Yuan, W. (2004). Regression models for unbalanced longitudinal ordinal data: computer software and a simulation. *Computer Methods and Programs in Biomedicine*, 75:195–200.

[52] Zeger, S. and Liang, K. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1):121–130.

[53] Ziegler, A. (2011). *Generalized estimating equations.* Springer-Verlag.

[54] Ziegler, A. and Gromping, U. (1998). The generalised estimating equations: a comparison of procedures available in commercial statistical software packages. *Biometrical Journal*, 40(3):245–260.

[55] Ziegler, A., Kastner, C., and Blettner, M. (1998). The generalised estimating equations: an annotated bibliography. *Biometrical Journal*, 40(2):115–139.

[56] Zorn, C. (2001). Generalized estimating equation models for correlated data: a review with applications. *American Journal of Political Science*, 45(2):470–490.