



Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R

Nick R. Parsons*, Matthew L. Costa, Juul Achten, Nigel Stallard

Warwick Medical School, University of Warwick, UK

ARTICLE INFO

Article history:

Available online 9 August 2008

ABSTRACT

The widely used proportional odds model is developed for correlated repeated ordinal score data, using a modified version of the generalized estimating equation (GEE) method for model fitting for a range of working correlation models. The algorithm developed estimates the correlation parameter, by minimizing the generalized variance of the regression parameters at each step of the fitting algorithm. Methods for parameter estimation are described for the widely used uniform and first-order autoregressive correlation models, for data potentially recorded at irregularly spaced time intervals. A full implementation of the algorithm (repolr) in the R statistical software package, that both tests the assumption of proportional odds and accommodates missing data, is described and applied to a clinical trial of post-operative treatment, after rupture of the Achilles tendon and a study of patient pain response after hip joint resurfacing.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Ordinal score variables, that have a clear hierarchical ordering, recorded from the same patient or experimental unit over time are common in clinical research studies. For instance a pain or discomfort score (mild, moderate or severe) may be recorded at a number of occasions, corresponding to routine assessments or visits to a clinic, or may be used to establish a patient's state at entry to a clinical trial and again at the conclusion, with differences between assessments attributed to the treatment effect. Ordinal score scales are often used to quantify symptoms or conditions that are difficult or impossible to assess in any other way, and are thus commonly used for evaluating the effectiveness of many surgical procedures (e.g. hip replacement or tendon repair) at a number of repeat occasions post-operatively; these two applications provide motivating examples for this paper.

Ordinal scores are common in health-related research, and many approaches have been described for developing regression models for efficient analysis of these data (Lall et al., 2002). In particular, the modelling of repeated ordinal scores is a widely studied statistical problem and an active area of research; Agresti and Natarajan (2001) provide a comprehensive review of available models and methods. A number of parametric and nonparametric methods have been proposed for analysis of repeated ordinal responses; the two most widely used approaches to modelling differ in their formulation based on either population-averaged or subject specific effects (Diggle et al., 2002). The latter models represent subject (or cluster) effects by a random effects term in the model; see for example Coull and Agresti (2000). The most widely used approach for population-averaged (marginal) models for repeated ordinal responses (e.g. Clayton (1992), Lipsitz et al. (1994), and Kenward et al. (1994)) is the generalized estimating equation (GEE) method originally proposed by Liang and Zeger (1986) for the proportional odds model (McCullagh, 1980); the model formulation follows directly from considering an ordinal score to be a continuous (unobserved) variable that is divided into a small number of categories, in an attempt to provide an objective evaluation of a quantity that would otherwise be impossible to measure directly. GEE methods provide reliable parameter

* Corresponding author. Tel.: +44 2476968627; fax: +44 2476968617.

E-mail address: nick.parsons@warwick.ac.uk (N.R. Parsons).

estimates in most situations, however, a number of alternative GEE methods have been suggested that overcome some of the pitfalls in estimation of the correlation parameters that occur in many of the most widely used methods (Chaganty, 1997; Parsons et al., 2006).

Commercial software for model fitting using the GEE method is now widely available (Horton and Lipsitz, 1999), although implementation for ordinal regression models is often difficult, even for an experienced user. In practice in many clinical trials, information in the data is often not used efficiently, and ordinal scores are treated as if they were continuous variables for the purposes of analysis (Forrest and Andersen, 1986; Lavalley and Felson, 2002; Jakobsson, 2004); the main reason for this is the lack of available, simple to use, statistical tools. In order to address this perceived problem we have generalized and extended the algorithm originally suggested by Parsons et al. (2006), which they show to be particularly robust and reliable, for model parameter estimation and analysis of typical clinical trials data and written functions for routine implementation and distribution in the freely available statistical software R (R Development Core Team, 2007). We extend the methodology, which was previously only available for the very limited case of complete ordinal scores recorded at evenly spaced time intervals with an assumed first-order autoregressive correlation model, in a number of ways that will facilitate more widespread use for analysis of data from clinical trials. To achieve this, we develop correlation models that accommodate positively correlated repeated ordinal scores within subjects recorded at any desired spacing of time intervals and, as data from clinical follow-up are in general incomplete due to patient drop-out, we modify the algorithm to allow missing data in either the response or explanatory variables, and also incorporate a formal test of the proportional odds assumption into the algorithm. We introduce the proportional odds model in Section 2.1, describe an algorithm for parameter estimation in Section 2.3 and give examples of correlation models appropriate for clinical trial data in Section 2.4. Section 3 describes R functions that implement the GEE method, and other available methods in R for modelling ordinal score data, and Section 4 demonstrates model fitting for two example sets of data; first a surgical trial of post-operative treatment after rupture of the Achilles tendon and second, a study of patient pain response after hip joint resurfacing.

2. Methods

2.1. Proportional odds model

Suppose in a clinical research study, that ordinal scores are recorded at each of T time points from N subjects (patients) using a scale with K categories. Let Y_{it} represent the score on the i th subject at the t th time, where the individual scores are made on an ordered categorical scale, where integers from 1 to K represent the score categories and K represents the optimum score category (for example, 1 = mild, 2 = moderate and 3 = severe). Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT})^T$ be a vector of ordinal scores for the i th subject at T time points and let \mathbf{x}_{it} be a matrix of observed explanatory variables (i.e. treatments, baseline covariates etc). The probability that an ordinal score Y_{it} falls within a particular score category can be related to the measured explanatory variables \mathbf{x}_{it} by a proportional odds model based on cumulative logits (McCullagh and Nelder, 1989, Chapter 5).

$$h(\mu_{itk}) = \beta_{0k} + \mathbf{x}_{it}^T \boldsymbol{\beta}. \quad (1)$$

Here, $\mu_{itk} = P(Y_{it} \leq k)$ is the probability that score Y_{it} is less than or equal to k , the parameters β_{0k} for $k = 1, \dots, K - 1$ are so-called cut-points ($-\infty < \beta_{01} < \dots < \beta_{0(K-1)} < \infty$), $\boldsymbol{\beta}$ is a vector of model regression parameters and $h(\cdot)$ is the logit function. The cumulative link model (Agresti, 2002) used in Eq. (1) effectively transforms the ordinal scale to a continuous scale based on the linear predictor $\mathbf{x}_{it}^T \boldsymbol{\beta}$, which proves to be a convenient model for relating the ordinal scores to the measured variates; the cut points β_{0k} define the class boundaries on the continuous scale of measurement defined by the linear predictor. The model is known as the proportional odds model, or proportional odds logistic regression model, because the ratio of the odds of the event $Y_{it} \leq k$ for any pair of sets of explanatory variables is independent of the choice of score category k . The applicability and intuitive assumptions behind the proportional odds model are the reasons for its popularity and widespread use. The assumption of proportional odds can be tested using the efficient score test suggested by Stiger et al. (1999). The cumulative logit model naturally results from an underlying latent variable with a logistic distribution, latent variables with assumed normal or extreme value distributions similarly lead to the probit or complementary log–log links, which provide practical alternatives to the proportional odds model (McCullagh and Nelder, 1989).

2.2. GEE

The most widely used method for parameter estimation for the proportional odds model when modelling correlated data, and the one described here, is the generalized estimating equation (GEE) approach (Hardin and Hilbe, 2002), first proposed by Liang and Zeger (1986). The method avoids strong assumptions about the distribution and the dependence structure of the repeated measures that would typically be required for a full likelihood model, by solving multivariate analogues of the quasi-likelihood estimating equations (Wedderburn, 1974). Resulting estimates are not maximum likelihood, but are asymptotically Normal and consistent (Liang and Zeger, 1986). Various options for implementation of GEE have been described elsewhere, but the most widely used method, and the one we adopt here, models the effects of the explanatory variables $\boldsymbol{\beta}$ and the correlation between repeated observations on an individual subject $\boldsymbol{\alpha}$, separately. The correlation

structure is characterized by a working correlation matrix \mathbf{R} that could be constructed using *a priori* assumptions about the likely correlation structure, but is usually calculated and updated at each step of an iterative fitting algorithm. For most clinical data analysis problems, inferences on the treatment effects are of prime importance and are estimated by solving an estimating equation $Q_\beta(\beta; \alpha) = 0$ for β , where α are treated as nuisance parameters. Estimates of α are often based on Pearson residuals, for example Lipsitz et al. (1994) give estimators for α for a range of common working correlation structures using this approach, however, problems such as non-convergence and estimates of α that are not sensible for the working correlation model have led some authors (Chaganty, 1997; Wang and Carey, 2003) to suggest alternative methods for estimating α . Parsons et al. (2006) propose a particularly robust method for estimating α that minimizes the logarithm of the determinant of the covariance matrix of the regression parameters at each step of a fitting algorithm, and show how it could be applied to modelling evenly spaced ordinal scores with an assumed first-order autocorrelation model for complete data. We generalize this methodology to accommodate unevenly spaced time intervals between scores, introduce an alternative uniform correlation model and a test of the proportional odds assumption, using code developed in the freely available statistical software R, to facilitate widespread use.

2.3. Parameter estimation

To form the cumulative binary variables necessary for fitting a repeated measures proportional odds model, the ordinal score for each subject, at each time point, is transformed into a new set of $K - 1$ variables; $Z_{itk} = 1$ if $Y_{it} \leq k$ or $Z_{itk} = 0$ if $Y_{it} > k$ for $k = 1, \dots, K - 1$. This allows the proportional odds model to be expressed as

$$\mathbf{E}[Z_i] = h^{-1}(\mathbf{X}_i\beta), \quad (2)$$

where \mathbf{X}_i is a complete design matrix, including cut points, \mathbf{Z}_i is a new binary response variable formed by grouping Z_{itk} for the i th subject and β is a vector of cut-point and regression parameters (Parsons et al., 2006). For given data matrices \mathbf{X}_i and coded response variables \mathbf{Z}_i for each subject, β can be estimated by iterative re-weighted least squares, by equating the generalized estimating equations $Q_\beta(\beta; \alpha)$ to zero where

$$Q_\beta(\beta; \alpha) = \sum_{i=1}^N \mathbf{D}_i^T \mathbf{W}(\alpha)_i^{-1} (\mathbf{Z}_i - \mu_i). \quad (3)$$

$\mathbf{D}_i = \mu'(\beta)_i$ is the first derivative of $\mu(\beta)_i$ with respect to β and $\mathbf{W}(\alpha)_i^{-1} = \mathbf{V}_i^{1/2} \mathbf{R}(\alpha)_i \mathbf{V}_i^{1/2}$. $\mathbf{R}(\alpha)_i$, a function of unknown correlation parameters α , is the matrix of correlations between the elements of \mathbf{Z}_i and $\mathbf{V}_i^{1/2}$ is a matrix containing the square roots of the variances of the elements of \mathbf{Z}_i , given by $\{\mu_{itk}(1 - \mu_{itk})\}^{1/2}$, along the leading diagonal. Under mild regularity conditions (Liang and Zeger, 1986), estimates of the regression parameters β are asymptotically multivariate normal with covariance matrix given by the so-called robust or sandwich estimator

$$\mathbf{V}_\beta(\alpha) = \left[\sum_{i=1}^N \mathbf{D}_i^T \mathbf{W}(\alpha)_i^{-1} \mathbf{D}_i \right]^{-1} \left[\sum_{i=1}^N \mathbf{D}_i^T \mathbf{W}(\alpha)_i^{-1} \text{cov}(\mathbf{Z}_i) \mathbf{W}(\alpha)_i^{-1} \mathbf{D}_i \right] \left[\sum_{i=1}^N \mathbf{D}_i^T \mathbf{W}(\alpha)_i^{-1} \mathbf{D}_i \right]^{-1}, \quad (4)$$

which provides a consistent estimator of the variances of the regression parameters, regardless of the specification of $\mathbf{W}(\alpha)_i$. In principle, Eq. (3) gives consistent estimates of the model parameters, even if the covariance structure of \mathbf{Z}_i is not correctly specified (Liang and Zeger, 1986). However, in practice, the efficiency of estimation of the parameter vector β and the reliability of inference can be improved if $\mathbf{R}(\alpha)_i$, the ‘working’ correlation matrix, is chosen to be as close to the true correlation matrix of \mathbf{Z}_i as possible.

For a single parameter α , an alternative to formulating and solving an additional estimating equation for the correlation model (Hardin and Hilbe, 2002) is to directly minimize $f(\alpha) = \log |\mathbf{V}_\beta(\alpha)|$, the logarithm of the generalised variance of β , with respect to α at each step of a model fitting algorithm. An updated estimate of the transformed correlation parameter ϕ , where $\phi = g^{-1}(\alpha)$ and $g(\cdot)$ is a correlation transformation function that ensures $\mathbf{R}(\alpha)_i$ remains positive definite during model fitting, after m iterations of the Newton–Raphson method using current estimates of the model parameters $\tilde{\beta}_m$, $\tilde{\alpha}_m$ and covariance matrix $\mathbf{V}_{\beta_m}(\alpha)$, is given by

$$\tilde{\phi}_{m+1} = \tilde{\phi}_m - \frac{f'(\tilde{\phi}_m)}{f''(\tilde{\phi}_m)}, \quad (5)$$

where $f'(\phi) = g'(\phi)f'(\alpha)$ and $f''(\phi) = g''(\phi)f'(\alpha) + g'(\phi)^2 f''(\alpha)$. Expressions for the first $f'(\alpha)$ and second $f''(\alpha)$ derivatives of $f(\alpha)$, with respect to α are given by Parsons et al. (2006) and derivatives of the correlation transform function $g(\cdot)$ (see Section 2.4) are straightforward to calculate. Fitting proceeds by using an initial estimate of ϕ and Eqs. (3) and (4) to estimate β and $\mathbf{V}_\beta(\alpha)$, and an updated estimate of ϕ is given by Eq. (5). The fitting procedure is iterated until convergence is achieved. Estimates of parameters β and ϕ will be obtained provided that Eq. (3) has a solution $\tilde{\beta}$ for every ϕ . A second-order approximation for $f(\tilde{\phi})$ at $\tilde{\phi}$, assuming an approximate Normal distribution for $\tilde{\phi}$ (Parsons et al., 2006), allows the standard error for the parameter $\hat{\phi}$ to be estimated by $s = \left(\sqrt{f''(\tilde{\phi})} \right)^{-1}$.

2.4. Correlation models

A natural model for the working correlation matrix for the set of vectors of observed ordinal scores on the same subject is $\mathbf{R}(\alpha) = \mathbf{C}(\alpha) \otimes \mathbf{S}$, where $\mathbf{C}(\alpha)$ is the $T \times T$ dimensional matrix of correlations between times and \mathbf{S} is the $(K-1) \times (K-1)$ dimensional matrix of correlations between the derived binary responses \mathbf{Z}_i within each time point. Assuming the same correlation matrix applies to every set of binary variables at every time point t , then the matrix of between score correlations is given by

$$\mathbf{S} = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1(K-1)} \\ \vdots & \ddots & \vdots \\ \rho_{(K-1)1} & \cdots & \rho_{(K-1)(K-1)} \end{bmatrix},$$

where for the proportional odds model, the expected correlation between binary variables Z_{itj} and Z_{itk} is given by $\rho_{jk} = \rho_{kj} = \{\exp(\beta_{0j} - \beta_{0k})\}^{1/2}$ where $j < k$ (Kenward et al., 1994). The simplest and most widely used models for the between time correlations are the independence, uniform and the first order autoregressive models (Diggle et al., 2002).

(i) Independence

Often the simplest choice for the correlation structure when modelling the within-subject between time correlations at time points t_i and t_j , is the independence model where $\mathbf{C}_{ij} = 1$ if $i = j$ and $\mathbf{C}_{ij} = 0$ otherwise, for example

$$\mathbf{C}(\alpha) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

(ii) Uniform

The next step up from independent errors is to equally correlated errors, where there is assumed to be a positive correlation α between any pair of measurements on the same subject $\mathbf{C}_{ij}(\alpha) = 1$ if $i = j$ and $\mathbf{C}_{ij}(\alpha) = \alpha$ otherwise, giving a correlation matrix

$$\mathbf{C}(\alpha) = \begin{bmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha \\ \alpha & \cdots & \alpha & 1 \end{bmatrix}.$$

The uniform correlation model in which a linear model for the mean response incorporates a random intercept term between subjects, leads directly to its interpretation as a random effects model with a random intercept for each subject.

(iii) First-order autoregressive (AR1)

Assuming a first-order autoregressive (AR1) model for α , the correlation matrix between vectors of observations at time points t_i and t_j becomes $\alpha^{|t_i - t_j|}$. Under these assumptions, and defining $d_{ij} = d_{ji} = |t_i - t_j|$, a natural model for the working correlation matrix for the set of vectors of observations on the same subject is $\mathbf{C}_{ij}(\alpha) = 1$ if $i = j$ and $\mathbf{C}_{ij}(\alpha) = \alpha^{d_{ij}}$ otherwise, for example

$$\mathbf{C}(\alpha) = \begin{bmatrix} 1 & \alpha^{d_{12}} & \cdots & \alpha^{d_{1T}} \\ \alpha^{d_{21}} & 1 & & \vdots \\ \vdots & & \ddots & \alpha^{d_{T-1T}} \\ \alpha^{d_{T1}} & \cdots & \alpha^{d_{TT-1}} & 1 \end{bmatrix}.$$

To ensure that $\mathbf{R}(\alpha)$ remains positive definite during model fitting for the uniform and first-order autoregressive models, α is constrained to lie in the interval $(0, 1)$ by making the transformation $\phi = g^{-1}(\alpha) = \log(\alpha) - \log(1 - \alpha)$ for $0 < \alpha < 1$. For model fitting the matrix $\mathbf{C}(\alpha)^{-1}$ and the first and second derivatives with respect to α are required; these are given in the Appendix. Typically, the GEE analysis is performed and the model robustness is assessed by the variation in magnitude of the regression parameter estimates using a range of models for $\mathbf{C}(\alpha)$.

3. Software implementation in R

R (R Development Core Team, 2007) functions to implement the above formulae for parameter estimation for the proportional odds models are available from the authors at <http://www.warwick.ac.uk/go/repolr>. The model fitting function `repolr` (repeated measures proportional odds logistic regression) has a number of required and optional arguments; the

Table 1

EuroQol activity scores for 48 patients after operative treatment for rupture of the Achilles tendon immediately after immobilisation (baseline), at six months and one year

Activity scores ^a			Number of patients	
Baseline	6 months	12 months	Treatment group (n = 22)	Control group (n = 26)
1	1	1	2	2
1	1	–	1	0
–	1	1	3	5
–	1	–	1	0
–	–	1	1	0
1	2	1	0	2
1	1	2	0	1
2	1	1	6	6
2	1	–	1	0
–	2	1	1	0
2	2	1	3	2
2	2	2	0	4
2	2	–	0	1
3	1	1	1	2
3	1	–	1	0
3	–	–	1	0
3	2	2	0	1

^a Scores are interpreted as 1 = no problem, 2 = some problem and 3 = inability to perform usual activities and missing scores are indicated by a hyphen.

correct syntax for implementation is described in detail at this site. The R function *repolr* fits the models described in Section 2, using all the available correlation options and gives estimated model parameters together with an efficient score statistic for the proportional odds assumption (Stiger et al., 1999). These functions provide a simple set of R procedures for analysis of repeated ordinal scores for the proportional odds model. The other option for model fitting using GEE in R is the function *ordgee*, which is part of the package *geepack* (Halekoh and Højsgaard, 2006), which produces a GEE fit of clustered ordinal data using the methods of Heagerty and Zeger (1996). This uses the same estimating equations as *repolr* for the mean regression parameters, but rather than parameterize the association between pairs of ordinal scores using correlation coefficients, it uses odds ratios to summarize the association using a single parameter for each pair of responses. The function *ordgee* allows independence, uniform, unstructured and user defined association models, but no obvious analogue of the widely used first order auto-regressive model is available using this methodology. As noted by Agresti and Natarajan (2001), methods for maximum likelihood fitting of marginal models are difficult to implement using standard methods. A comprehensive review of these methods is beyond the scope of this paper, but a number of methods is available as R libraries. The package *drm* uses a likelihood-based approach for the analysis of clustered ordinal data (Ekholm et al., 2003; Jokinen, 2006), that puts equal emphasis on the regression model and on dependence ratios that characterise the association between responses. Another available R package *mprobit* allows maximum likelihood fitting of the ordinal probit model for AR1, uniform and unstructured correlation models. Lindsey and Kaufmann (2004) offer a flexible approach to the conditional modelling of ordinal longitudinal responses that is available using the ordinal package.

4. Examples

Model fitting and analysis using *repolr* are illustrated for two examples from clinical trials. The first is a randomized controlled trial on the effectiveness of immediate weight-bearing mobilisation after rupture of the Achilles tendon. The second uses data collected from a prospective study of metal-on-metal hip resurfacing implants.

4.1. Achilles tendon rupture

A randomized clinical trial was performed in order to assess the potential benefits of immediate weight-bearing mobilisation after rupture of the Achilles tendon (Costa et al., 2006). Data were collected from 48 patients who decided on operative treatment, and were randomised to either a treatment group (immediate mobilisation in a carbon-fibre orthosis with three 1.5 cm heel raises) or a control group (traditional plaster cast immobilisation). Each patient was assessed at baseline, and reviewed at six months and one year using the EuroQol questionnaire (Brooks, 1996); a widely used quality of life measure where patients are asked to assess their own health using five factors – mobility, self-care, usual activity, pain/discomfort and anxiety/depression – on a 3-point ordinal scale for each factor. For this study, the key questionnaire factor of interest was a patient's ability to undertake their 'usual activities' during the first year after surgery; this was scored by each patient as either no problem (1), some problem (2) or an inability to perform (3) usual activity (e.g. work, leisure, housework etc.). The resulting data are summarised in Table 1 by treatment and control groups and are available at <http://www.warwick.ac.uk/go/repolr>; example code to fit all the models described here is also available from this site.

A naïve cross-sectional analysis of the data showed that activity scores declined, indicating an improvement in patient activity, over time for both control (baseline = 1.90, 6 months = 1.38 and 12 months = 1.24) and treatment (baseline =

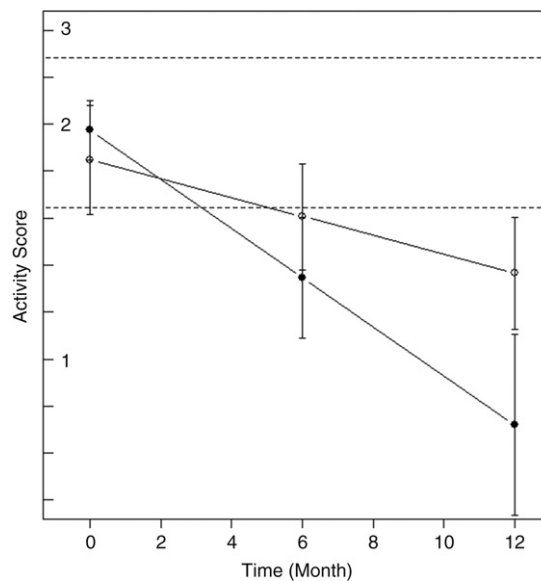


Fig. 1. Predicted mean activity scores on the scale of the linear predictor, error bars (± 1 standard errors of predicted mean) and divisions between score categories (---), estimated by cut-points for the control (○) and treatment (●) groups post surgery after rupture of the Achilles tendon.

Table 2

Parameter estimates and standard errors (s.e.) for AR1, uniform and independence correlation models, for regression models for post-operative treatment after rupture of the Achilles tendon

Parameter	Correlation model					
	AR1		Uniform		Independence	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
β_{01}	−4.783	1.119	−4.794	1.105	−4.828	1.132
β_{02}	−1.597	1.071	−1.610	1.056	−1.589	1.087
Treat	2.552	1.280	2.586	1.266	2.547	1.298
Time	3.131	0.641	3.139	0.631	3.164	0.645
Treat \times Time	−1.923	0.731	−1.938	0.721	−1.932	0.740
ϕ	−2.207	5.298	−2.454	4.616	–	–

2.00, 6 months = 1.20 and 12 months = 1.00) groups, with some evidence to suggest that scores returned to normal levels more rapidly for the treatment group than the control group. The three category ordinal response (activity score) was modelled with *repolr* using each of the available correlation structures, investigating the interaction of a linear model for the equally spaced follow-up occasions (time variable) with the treatment factor (control group or treatment group). The algorithm converged in five iterations for all correlation models giving the estimates of model parameters shown in Table 2. The score test statistic was not significant for the AR1 model (2.965 on 3 d.f.; p -value 0.397), indicating that the proportional odds assumption can not be rejected; similar results were also obtained for the other correlation models. The sign and magnitude of the regression coefficients varied little between the correlation models, suggesting that these parameters were robust to covariance specification. The estimated correlation parameters were small for the uniform ($\alpha = 0.079$) and AR1 models ($\alpha = 0.099$). The parameterisation is such that those attributes that increase the activity score will have negative regression coefficients (i.e. make patients worse), whereas those that decrease the score will have positive regression coefficients. The positive regression coefficient for post-surgery time (Z -value = 4.89), suggested that the scores decreased significantly over time, indicating an improvement in activity, as patients recovered after surgery, and the negative coefficient for the interaction term (Z -value = −2.632), suggested that the rate of improvement was significantly more rapid for the treatment group than the control group. The treatment group returned to normal activity more rapidly than the control group. A concise summary of the analysis is provided by plotting predicted activity scores and cut-points on the scale of the linear predictor (Fig. 1); methods have also been suggested for presenting results of ordinal regression analysis on the original score scale (Hannah and Quigley, 1996). This illustrates the decline and difference in activity scores between treatment groups and the relationship with cut-point divisions between score categories.

Table 1 shows that a sizeable proportion (16 out of 48) of the patients did not provide complete scores at all three occasions. This is not untypical for data collected from the clinical follow-up of patients, which are often incomplete due to patient drop-out. The *repolr* algorithm estimates the regression parameters using the *gee* package, which accommodates missing data by omitting individual data records where data is missing in either the response or explanatory variables. The original algorithm suggested by Parsons et al. (2006) for estimating α which did not accommodate missing data has been

Table 3

Hip function pain scores for 58 patients after hip resurfacing surgery at one year, two years and five years

Pain scores ^a			Number of patients	
1 year	2 years	5 years	Female (n = 21)	Male (n = 37)
1	1	1	7	18
1	1	2	0	3
1	1	3	1	0
1	2	1	2	1
1	2	2	0	1
1	2	3	1	0
1	2	4	1	0
1	3	3	0	1
1	4	3	0	1
1	4	4	1	0
2	1	1	0	1
2	1	2	1	3
2	1	3	1	0
2	2	2	1	1
2	2	3	0	1
2	2	4	1	0
2	3	3	1	0
3	1	2	0	1
3	1	4	0	2
3	2	1	1	0
3	2	2	0	1
3	3	2	1	0
3	3	4	1	2

^a Scores are interpreted as 1 = none, 2 = slight, 3 = mild and 4 = moderate or marked pain.

modified using a complete case analysis. For the modified algorithm implemented in *repolr*, if missing data are supplied for r subjects then data for all r subjects is excluded and estimation of α is based solely on complete data available from $N - r$ subjects (complete case analysis); see Jansen et al. (2006) for a review of methods for incomplete data for longitudinal clinical trial data. This approach will provide consistent estimates of the model parameters, provided that the data are assumed missing completely at random (MCAR). The validity of inferences drawn from the models on the incomplete Achilles tendon data is dependent on whether the mechanism generating the missing data is ignorable. Following Hardin and Hilbe (2002), a test of the proportion of missing activity scores in the treatment and control groups at each occasion provided no evidence to reject the hypothesis that scores were MCAR (p -values; 0.752, 0.398 and 0.125). Often the assumption of MCAR is unrealistic for repeated measurement data, and resulting model parameters may be biased; for a full discussion of available models and methods for missing data see for example Schafer and Graham (2002).

4.2. Hip resurfacing

Total hip replacement surgery is a widely used procedure to alleviate pain and debilitation caused by osteoarthritis, rheumatoid arthritis, fractures and other hip related problems that has been shown to give excellent results in older patients with limited functional demand. However, in younger more active patients there is a higher rate of long-term failure, due to the aseptic loosening associated with polyethylene wear. This problem and improvements in engineering in recent years have led to the re-emergence of alternative approaches to hip arthroplasty, including the use of large-diameter metal-on-metal hip resurfacing. The long-term results of hip resurfacing arthroplasty remain unclear, and much current research is aimed at assessing the failure rate and prognosis after surgery for this procedure. Costa et al. (in press) report on 249 consecutive metal-on-metal resurfacings of the hip implanted between September 1995 and December 2002. Harris hip scores (Harris, 1969) were collected at one, two and five years post-operatively from a subset of 58 patients who underwent hip resurfacing. The Harris hip score (HHS) has a number of components that measure hip function; the most important being a pain score coded for the purpose of this analysis as none (1), slight (2), mild (3) and moderate or marked (4) pain in the hip joint. The resulting data are summarised in Table 3 by sex and are also available at <http://www.warwick.ac.uk/go/repolr>, as is code for model fitting. For this study the key questions of interest were (i) whether pain scores increased over the five year period studied, indicating wear on the resurfacing implant, and (ii) were there significant differences in pain scores between sexes.

The four category ordinal response (pain score) was modelled with *repolr* using each of the available correlation structures, investigating the interaction of the assessment occasions with a sex factor (male and female). There was evidence for a significant correlation parameter for both the uniform ($\alpha = 0.331$) and the AR1 models ($\alpha = 0.714$), using the appropriate time spacing of one, two and five years for the AR1 model. Based on the asymptotic Normality of $\tilde{\phi}$, a Wald-type two sided confidence interval showed that there was good evidence for a strong positive correlation between repeated pain scores for the AR1 model; the 95% CI for $\tilde{\alpha}$ was (0.35, 0.92), obtained by transforming the interval for $\tilde{\phi}$ using the correlation transformation function $g(\cdot)$. The score test statistic was not significant for the AR1 model (8.768 on 10 d.f.; p -value 0.554),

Table 4

Parameter estimates and standard errors (s.e.) for the regression model for pain score after resurfacing hip surgery

Parameter	Estimate	s.e.	Z-value
β_{01}	0.574	0.450	1.274
β_{02}	1.635	0.441	3.707
β_{03}	2.844	0.463	6.140
Sex (male)	0.176	0.561	0.315
Occasion 2 (year 2)	−0.547	0.428	−1.279
Occasion 5 (year 5)	−0.971	0.475	−2.044
Sex × Occasion 2	0.940	0.613	1.533
Sex × Occasion 5	0.419	0.544	0.771
ϕ	0.914	0.792	–

indicating that the proportional odds assumption cannot be rejected. The algorithm converged in eight iterations for the AR1 model giving the parameter estimates shown in Table 4. The significant negative regression coefficient for year 5 after surgery (Z -value = -2.044) suggested that the pain scores were significantly higher at this occasion, indicating a clear deterioration in hip function after this period of time substantiating the observed increase in hip failure rate observed by Costa et al. (in press). There was no evidence to indicate a significant increase (Z -value = -1.279) in a patient's pain score at the second year after surgery. Also, there was no evidence to indicate a significant difference in pain scores between sexes (Z -value = 0.315), to support the observed higher failure rate reported for females by Costa et al. (in press), and no suggestion that the observed increase in pain scores at year 5 differed between sexes.

5. Discussion

We have presented a methodology for the analysis of repeated ordinal score data using the proportional odds logistic regression model, that allows successive observations on the same subject to be positively correlated over time. The GEE methodology used for parameter estimation, originally proposed for use in a very different application (Parsons et al., 2006), has been extended, modified and implemented for routine use in the R software environment (R Development Core Team, 2007). The R package *repolr* provides a complete set of procedures for routine analysis of repeated ordinal score data using the proportional odds model that accommodates missing values and tests the proportional odds assumption for widely used correlation models. Although the algorithm used for parameter estimation has been shown to be efficient and more robust than other GEE methods for a limited range of data (Parsons et al., 2006), a comprehensive review of all available procedures for fitting of ordinal regression models would be required to fully understand the relative merits of one approach in comparison to another.

Data from longitudinal studies in general, and from clinical trials in particular, are prone to incompleteness. The complete case analysis used in *repolr* will provide consistent estimates of the model parameters, provided that the data are assumed missing completely at random (MCAR). As estimation of the correlation parameter is based on a complete case analysis, if no single subject or experimental unit in a data set contains a complete series of scores at all occasions the algorithm will fail to make progress; given that in most clinical trials patients are followed-up at no more than 2–3 occasions after treatment, we do not regard this as a major problem.

Generalized estimating equations provide a useful method to avoid the computational complexity of a full likelihood analysis, when the regression parameters representing treatments, baseline covariates and time effects are the prime focus of an analysis. As our interest is mainly in estimating the regression parameters, we regard the association parameters as a nuisance, and choose to use simple, but robust, models using a single parameter to model the between time correlations for each subject and assume that the same correlation matrix applies to every set of binary variables at every time point. If association parameters are of scientific interest then the *geepack* package (Halekoh and Højsgaard, 2006), also available in R, implements the approach to ordinal regression proposed by Heagerty and Zeger (1996). The hip resurfacing model showed a strong sensitivity to α , that was not apparent for the Achilles tendon rupture model. However, as consistency is based on large sample properties, estimates of α for the latter example are probably not reliable due to the small sample size used for estimation of the correlation parameter due to missing data; a not uncommon problem for much data from clinical trials. Function *repolr* also provides an option to fix the correlation parameter at its initial value, and not to update estimates, during model fitting that provides a useful tool for assessing the sensitivity of individual regression parameters to α . The desirable model robustness properties of $V_{\beta}(\alpha)$ (Kauermann and Carroll, 2001) explain its widespread use, often in preference to the model based variance estimate also available when using *repolr*. However, it has been noted in the literature that the robust estimator may not work well with small samples (e.g. Drum and McCullagh (1993)), a point other authors illustrate for binary data using simulation studies (e.g. Emrich and Piedmonte (1992)), and has lead some to suggest modification to $V_{\beta}(\alpha)$ to overcome these problems (Pan and Wall, 2002). The approach used here could in principle be generalized to accommodate models with more than one correlation parameter, although this seems at odds with our stated aim to provide relatively simple models for α . The Newton–Raphson method will converge to the correct value for α provided the function $f(\alpha)$ is concave, although occasionally the algorithm may converge to a local minimum or cycle between two points. This can usually be overcome by starting the algorithm at a different initial value for α , ideally if this is nearer to the global minimum it will improve convergence.

The repeated measures ordinal regression algorithm, described in Section 2, can be implemented in R using the function `repolr`. The algorithm has good convergence properties, and is fast and generally not sensitive to the initial value for α ; when the algorithm does fail to converge, even after adjusting the starting value for α , then this is often because the chosen model is not suitable for the observed data, and the user is advised to try a different model formulation. As the algorithm calls the GEE solver (`gee`) at each iteration to update estimates of the model regression parameters, there will on occasion be warnings of a lack of convergence at the early stages of the optimisation reported by `gee`, which can safely be ignored if the full algorithm converges successfully. The algorithm monitors convergence using the average relative change in the regression parameters at each iteration, by default terminating the procedure when this drops below 0.1%. However, to confirm that a true minimum has been determined for the transformed correlation parameter ϕ at convergence we also require that $f'(\phi) \approx 0$ and $f''(\phi) > 0$; these are monitored and reported by `repolr` at each iteration. Model diagnostics are available using the standard tools available within R and hypothesis tests (e.g. Wald tests) to compare nested models can be constructed using outputs of β and $V_\beta(\alpha)$ (Rotnitzky and Jewell, 1990). However, as for all logistic regression models where data are sparse (i.e. there is insufficient data to justify large sample results) goodness-of-fit tests are problematic; Barnhart and Williamson (1998) and Lee and Qaqish (2004) discuss goodness-of-fit methods for GEE.

Acknowledgements

The authors would like to thank two anonymous reviewers for their useful comments on earlier versions of this paper.

Appendix. Expressions for $C(\alpha)^{-1}$, $C'(\alpha)^{-1}$ and $C''(\alpha)^{-1}$

(i) For the uniform correlation model $C(\alpha)^{-1} = \begin{bmatrix} a & b & \cdots & b \\ b & \ddots & & \\ \vdots & \ddots & \ddots & \\ b & \cdots & b & a \end{bmatrix}$, where $a(\alpha) = \frac{-(1+(n-2)\alpha)}{(n-1)(\alpha-1)(\alpha+\frac{1}{n-1})}$ and $b(\alpha) = \frac{\alpha}{(n-1)(\alpha-1)(\alpha+\frac{1}{n-1})}$; expressions for $C'(\alpha)^{-1}$ and $C''(\alpha)^{-1}$ follow from the derivatives $a'(\alpha)$, $a''(\alpha)$, $b'(\alpha)$ and $b''(\alpha)$.

(ii) For the AR1 correlation model $C(\alpha)^{-1} = \begin{bmatrix} a_2 & c_2 & 0 & \cdots & 0 \\ c_2 & b_3 & \ddots & & \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & b_T & c_2 \\ 0 & \cdots & 0 & c_2 & a_T \end{bmatrix}$, where $a_i(\alpha) = (1 - \alpha^{2d_{ii-1}})^{-1}$, $b_i(\alpha) = a_{i-1}(\alpha)a_i(\alpha)(1 - \alpha^{2d_{ii-2}})$ and $c_i(\alpha) = -\alpha^{d_{ii-1}}a_i(\alpha)$ for $t = 1, \dots, T$; expressions for $C'(\alpha)^{-1}$ and $C''(\alpha)^{-1}$ follow from the derivatives $a'_i(\alpha)$, $a''_i(\alpha)$, $b'_i(\alpha)$, $b''_i(\alpha)$, $c'_i(\alpha)$ and $c''_i(\alpha)$.

References

- Agresti, A., 2002. Categorical Data Analysis, 2nd ed. Wiley, New York.
- Agresti, A., Natarajan, R., 2001. Modeling clustered ordered categorical data: A survey. *International Statistical Review* 69, 345–371.
- Barnhart, H.X., Williamson, J.M., 1998. Goodness-of-fit tests for GEE modeling with binary responses. *Biometrics* 54, 720–729.
- Brooks, R., 1996. EuroQol: The current state of play. *Health Policy* 37, 53–72.
- Chaganty, N.R., 1997. An alternative approach to the analysis of longitudinal data via generalized estimating equations. *Journal of Statistical Planning and Inference* 63, 39–54.
- Clayton, D., 1992. Repeated ordinal measurements: A generalized estimating equations approach. Technical Report, MRC Biostatistics Unit, Cambridge, UK.
- Costa, M.L., MacMillan, K., Halliday, D., Chester, R., Shepstone, L., Robinson, A.H.N., Donell, S.T., 2006. Randomised controlled trials of immediate weight-bearing mobilisation for rupture of the tendon Achilles. *Journal of Bone and Joint Surgery (British)* 88-B, 69–77.
- Costa, M.L., O'Brien, M., Prasad, A., Parsons, N., Achten, J., Pereira, G., Prakash, U., Foguet, P., Griffin, D., Krikler, S.J., 2008. Resurfacing arthroplasty of the hip: Minimum 5 year results from an independent centre. *Journal of Bone and Joint Surgery (British)* (in press).
- Coull, B.A., Agresti, A., 2000. Random effects modelling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics* 56, 73–80.
- Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L., 2002. Analysis of Longitudinal Data. Oxford University Press, Oxford.
- Drum, M., McCullagh, P., 1993. Comment. *Statistical Science* 8, 300–301.
- Ekholm, A., Jokinen, J., McDonald, J.W., Smith, P.W.F., 2003. Joint regression and association modeling of longitudinal ordinal data. *Biometrics* 59, 795–803.
- Emrich, L.J., 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.
- Forrest, M., Andersen, B., 1986. Ordinal scale and statistics in medical research. *British Medical Journal* 292, 537–538.
- Halekoh, U., Højsgaard, S., 2006. The R package `geepack` for generalized estimating equations. *Journal of Statistical Software* 15, 2.
- Hannah, M., Quigley, P., 1996. Presentation of ordinal regression analysis on the original scale. *Biometrics* 52, 771–775.
- Hardin, J.W., Hilbe, J.M., 2002. Generalized Estimating Equations. Chapman and Hall, New York.
- Harris, W.H., 1969. Traumatic arthritis of the hip after dislocation and acetabular fractures: Treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *Journal of Bone Joint Surgery* 51-A, 737–755.
- Heagerty, P.J., Zeger, S.L., 1996. Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association* 91, 1024–1036.
- Horton, N.J., Lipsitz, S.R., 1999. Review of software to fit generalized estimating equation regression models. *The American Statistician* 53, 160–169.
- Jakobsson, U., 2004. Statistical presentation and analysis of ordinal data in nursing research. *Scandinavian Journal of Caring Sciences* 18, 437–440.

- Jansen, I., Beunckens, C., Molenberghs, G., Verbeke, G., Mallinckrodt, C., 2006. Analyzing incomplete discrete longitudinal clinical trial data. *Statistical Science* 21, 52–69.
- Jokinen, J., 2006. Fast estimation algorithm for likelihood-based analysis of repeated categorical responses. *Computational Statistics and Data Analysis* 51, 1509–1522.
- Kauermann, G., Carroll, R.J., 2001. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 96, 1387–1396.
- Kenward, M.G., Lesaffre, E., 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, 945–954.
- Lall, R., Campbell, M.J., Walters, S.J., Morgan, K., 2002. A review of ordinal regression models applied to health-related quality of life assessments. *Statistical Methods in Medical Research* 11, 49–67.
- Lavalley, M.P., Felson, D.T., 2002. Statistical presentation and analysis of ordered categorical outcome data in rheumatology journals. *Arthritis and Rheumatism* 47, 255–259.
- Lee, J.-H., Qaqish, B.F., 2004. Modified GEE and goodness of the marginal fit (GOMF) test with correlated binary responses for contingency tables. *Biometrical Journal* 46, 675–686.
- Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22.
- Lindsey, P.J., Kaufmann, J., 2004. Analysis of a longitudinal ordinal response clinical trial using dynamic models. *Applied Statistics* 53, 523–537.
- Lipsitz, S.R., Kim, K., Zhao, L., 1994. Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine* 13, 1149–1163.
- McCullagh, P., 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society Series B* 42, 109–142.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.
- Pan, W., Wall, M.M., 2002. Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. *Statistics in Medicine* 21, 1429–1441.
- Parsons, N.R., Edmondson, R.N., Gilmour, S.G., 2006. A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research. *Applied Statistics* 55, 507–524.
- R Development Core Team., 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.
- Rotnitzky, A., Jewell, N.P., 1990. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* 77, 485–497.
- Schafer, J.L., Graham, J.W., 2002. Missing data: Our view of the state of the art. *Psychological Methods* 7, 147–177.
- Stiger, T.R., Barnhart, H.X., Williamson, J.M., 1999. Testing proportionality in the proportional odds model fitted with GEE. *Statistics in Medicine* 18, 1419–1433.
- Wang, Y.-G., Carey, V., 2003. Working correlation structure misspecification, estimation and covariate design: Implications for generalized estimating equations performance. *Biometrika* 90, 29–41.
- Wedderburn, R.W.M., 1974. Quasi-likelihood functions, generalized linear models and the gaussian method. *Biometrika* 61, 439–447.