Random-effect models for ordinal responses: application to self-reported disability among older persons.

Isabelle Carrière, Jean Bouyer

Rev Epidemiol Sante Publique 54 (2006) 61-72

Random-effect models for ordinal responses: Application to self-reported disability among older persons.

I. Carrière[1] and J. Bouyer[2]

[1]INSERM Unité 500, 39 avenue Charles Flahault, 34093 MONTPELLIER, FRANCE

[2] INSERM Unité 569, 82 rue du Général Leclerc, 94276 LE KREMLIN BICETRE, France ; INED ; IFR69

Corresponding author:

Isabelle Carrière, INSERM Unité 500, 39 avenue Charles Flahault, 34093 MONTPELLIER, France

Email : carriere@montp.inserm.fr

Tel : 04 99 23 23 72

Fax: 04 99 61 45 79

Short title: Random-effect models for ordinal data

Abstract

Background

Longitudinal studies with ordinal repeated outcomes are now widespread in epidemiology and clinical research. The statistical analysis of these studies combines two difficulties: the choice of the best ordinal model and taking into account correlations for within-subject responses.

Methods

Random-effect models are of particular value in this context and we propose here a fitting strategy. The various ordinal models extended to the case of repeated responses are detailed. We explain how the choice of model constrains the random effect structure. Model selection criteria and goodness-of-fit measures are also presented.

These issues are dealt with by the mean of the example of self-reported disability in older women assessed annually over a period of seven years.

Results

The proportionality of the odds ratios was validated for the covariables "age" and "gait speed". In contrast the impact of the covariable "pain" differs according to the levels of disability. The restricted partial proportional odds model was found to have a goodness of fit equivalent to the full generalized ordered logit model while the stereotype model appeared to give poorer fit.

Conclusions

The random-effects models presented in this paper allow taking into account the ordinal nature of the outcome in longitudinal studies. Furthermore the impact of the risk factors can be modeled according to the response levels. This approach can be useful for a better understanding of complex processes of evolution.

Longitudinal studies. Random-effects models. Ordinal response. Disability evaluation

Résumé

Position du problème

Les études longitudinales avec des réponses ordinales répétées sont maintenant fréquentes en épidémiologie et recherche clinique. L'analyse statistique de ces études comporte deux écueils : le choix du meilleur modèle ordinal et la prise en compte des corrélations intra-sujets des réponses.

Méthodes

Les modèles à effets aléatoires sont particulièrement bien adaptés dans ce contexte et nous proposons dans cet article une stratégie de modélisation. Les différents modèles ordinaux étendus au cas des réponses répétées sont détaillés. Nous expliquons comment le choix du modèle impose des contraintes sur la structure des effets aléatoires. Nous présentons aussi des critères de sélection et des mesures d'adéquation des modèles aux données.

Ces questions sont illustrées par un exemple d'étude de l'incapacité mesurées chaque année, chez des femmes âgées, sur une période de sept ans.

Resultats

La proportionnalité des odds ratios a été validée pour les covariables "âge" et "vitesse de marche". Par contre l'impact de la covariable "douleur" est différent selon le niveau d'incapacité. L'adéquation aux données du modèle restreint à odds partiellement proportionnels était comparable à celle du modèle logistique ordonné complet tandis que celle du modèle stéréotype était moins bonne.

Conclusion

Les modèles à effets aléatoires présentés dans cet article permettent d'analyser les études longitudinales en prenant en compte le caractère ordinal de la réponse. De plus, l'impact des facteurs de risque peut être modélisé selon les niveaux de la variable réponse. Cette approche peut être utile pour mieux comprendre les profils d'évolution complexes.

Etudes longitudinales. Modèles à effets aléatoires. Réponse ordinale. Evaluation de l'incapacité.

Background

In epidemiology and clinical research, dimensions such as quality of life, pain, fatigue, depression, cognitive impairment, mood and physical disability are increasingly studied as outcomes of interest. In elderly people, for example, measurements of functional limitation are of key importance. In recent years, a large number of scales have been developed for the evaluation of these dimensions, most of them being ordinal indicators [1-4]. Moreover, the study of the succession of stages of disability, together with analyses of the risk factors for future deterioration is a major concern in gerontology. For this explanatory research, cross-sectional studies are inappropriate due to biases such as selective removal or differences in the proportions of impaired individuals between age groups, which may also be confounded with changes over generations [5]. Therefore, studies on risk factors for disability among elderly people require statistical tools able to take into account at the same time the ordinal nature of the outcome variable, the correlation between repeated observations for a subject and the time-dependence of some covariables (the covariables may vary over time for a subject). These issues may be addressed by means of random-effect models. However, few studies have used this type of model for the analysis of ordinal responses and, indeed, their application raises several problems.

In this paper, we describe a strategy for selecting and fitting multivariate random-effect models to longitudinal ordinal responses. We firstly present the models and their constraints coming from both the random-effect and the ordinal aspects of the models. Then, we provide tools for comparing their goodness of fit. Finally, this strategy is applied and the models are compared in an epidemiological study of changes over time of a disability indicator among a cohort of elderly women.

Methods

In longitudinal studies, each subject is assessed at different time points and these intra-individual responses are correlated. These correlations between individual responses must be accounted for by appropriate analysis methods [6] such as the generalized linear mixed model (GLMM), also known as the multilevel model or the conditional model [7]. The use of random effects in linear models for normal responses is well established. By contrast, random effects have only recently been incorporated into models for categorical data, due to complexities of implementation.

A basic characteristic of these models is the introduction of random subject effects into the linear predictor, which has led to random-effects models being described as "subject-specified models" [8]. The random part of the linear covariable combination describes and explains the structure of the longitudinal correlation and indicates the magnitude of between-individual heterogeneity.

Let Y_{ij} denote an outcome corresponding to the jth response (i.e. the response at time j) of the ith subject (i=1 to N). Let $\mathbf{X_{ij}}$ be a design matrix of time dependent covariates and $\mathbf{Z_{ij}}$ a submatrix of $\mathbf{X_{ij}}$. Let $\mathbf{U_i}$ denote the vector of random effect values for the subject i. Conditional on $\mathbf{U_i}$, a GLMM resembles an ordinary GLM. Let μ_{ij}=E(Y_{ij}/$\mathbf{U_i}$) denote the mean of the conditional distribution of Y_{ij} given $\mathbf{U_i}$. The linear predictor for a GLMM therefore takes the form:

$$g(\mu_{ij}) = \mathbf{X_{ij}'}\boldsymbol{\beta} + \mathbf{Z_{ij}'}\mathbf{U_i}$$

where g(.) is a link function and $\mathbf{U_i}$ is assumed to be independent of the covariates and normally distributed with mean zero and unknown variance.

For the fixed parameters β, estimates are obtained conditionally on random effects distribution. The likelihood function has no closed-form solution and various estimation procedures have been proposed to overcome this difficulty. The most attractive of these procedures directly maximizes an approximate likelihood obtained by numerical integration and is implemented for instance, in the SAS NLMIXED procedure. The likelihood is the integral, over the unobserved random effects, of the joint distribution of the data and the random effects. With non-linear models, computational intensive integration methods, such as the Gauss-Hermite quadrature [9], are required to evaluate this likelihood. Having approximated the likelihood, standard maximization methods, such as the Newton-Raphson method, are used to obtain the estimates.

Several models have been proposed for the analysis of ordinal responses in the case of non correlated data [10, 11]. These models can be extended to the case of repeated responses. However, the choice of the model imposes constraints on the random effect structure. Furthermore, in practice, it is difficult to introduce more than a few random effects with this estimation procedure. We present here several types of ordinal models with only a random intercept.

The proportional odds mixed models

Let Y_{ij} now denote an ordinal outcome with K categories.

The cumulative logit model was the first model proposed for ordinal responses. The ordinal nature of the response was taken into account by considering the cumulative probabilities $P(Y_{ij} \geq c)$, ordered as follows $1=P(Y_{ij} \geq 0) \geq P(Y_{ij} \geq 1) \geq \ldots P(Y_{ij} \geq c) \geq \ldots P(Y_{ij} \geq K)$.

The model can be represented as follows:

$$\log\left(\frac{P(Y_{ij} \geq c \mid \mathbf{X_{ij}}, u_i)}{P(Y_{ij} < c \mid \mathbf{X_{ij}}, u_i)}\right) = \alpha_c + \mathbf{X_{ij}'}\boldsymbol{\beta} + u_i \qquad (1)$$

where u_i is the random effect specific to subject i and α_c is the intercept for category c. The vector of the fixed parameters ($\boldsymbol{\beta}$) represents the log-odds ratios of the grouped categories superior to the cut-off (c) compared to the categories inferior to c. In this model, the random effect (u_i) and the fixed effects ($\boldsymbol{\beta}$) are independent of the cut-off (c). The assumption that $\boldsymbol{\beta}$ is independent of the response level implies that the log-odds ratio for a given covariable is the same for all levels of the response. This led McCullagh to refer to this model as the "proportional odds model" [12].

The consideration of cumulative properties confers several useful properties to the model: (1) The proportional odds model is invariant if the codes for the response **Y** are reversed, resulting only in a change in the sign of the regression parameters, (2) the proportional odds model is invariant if the categories of the ordinal response are collapsed [10]. The parameter vector, $\boldsymbol{\beta}$, remains unchanged if categories are combined, only the intercepts, α_c, are affected.

The proportional odds model is often presented as designed for ordinal response categories monotonically related to an underlying continuous latent variable $\boldsymbol{\eta}$. The relationship between **Y** and $\boldsymbol{\eta}$ is such that the parameters α_c are the cut-off points on the continuum $\boldsymbol{\eta}$ and that Y_{ij} equals c if η_{ij} lies between α_c and α_{c+1}. However, as pointed out by McCullagh [13], the existence of $\boldsymbol{\eta}$ is usually unverifiable in practice.

However, the use of a random effect, u_i, independent of response category is based on the notion that a unique unknown continuous phenomenon underlies the ordinal response. It is possible to consider a model including differential random effects, u_{ic}, by category of **Y**. However, estimation is more complicated in this extended model because

constraints must be added to ensure that the intercept ordering is not violated. For a given response category, c, the intercept is then: α_c+u_{ic} and therefore it is necessary that: $\alpha_1+u_{i1} \geq \alpha_2+u_{i2} \geq \ldots \geq \alpha_K+u_{iK}$.

The proportional odds assumption ($\boldsymbol{\beta}$ independent of the response level c) may be too strict and should, in any case, be tested. For this purpose the proportional odds model may be compared to more general ordinal models that permit the effects, $\boldsymbol{\beta}$, to vary according to the categories of $\mathbf{Y}$. The most general model of this type is called the generalized ordered logit model and can be expressed as follows:

$$\log\left(\frac{P(Y_{ij} \geq c \mid \mathbf{X_{ij}}, u_i)}{P(Y_{ij} < c \mid \mathbf{X_{ij}}, u_i)}\right) = \alpha_c + \mathbf{X_{ij}^{'}}\boldsymbol{\beta}_\mathbf{c} + u_i \qquad (2)$$

In the proportional odds model (1), the ordered nature of the response is expressed exclusively by the mean of the ordered intercepts, α_c. In contrast, in model (2), the covariates are involved in this ordering [14].

The hierarchy of cumulative probabilities implies the same ordering for the logits: logit($P(Y_{ij} \geq 1)$) $\geq \ldots$ logit($P(Y_{ij} \geq c)$) $\geq \ldots$logit($P(Y_{ij} \geq K\text{-}1)$). Thus, the linear predictors must also be identically ordered. For a single covariable, and as the random effect is thought to be independent of the response level, this relationship becomes: $\alpha_1 + \beta_1 x \geq \alpha_2 + \beta_2 x \geq \ldots \geq \alpha_{K-1} + \beta_{K-1} x$. These K-1 regression lines are not necessarily parallel and they may cross for some values of x, resulting in violation of the previous hierarchy, although this crossing of regression lines generally occurs outside the range of observed data. It may nonetheless be problematic if extrapolation of the model beyond the range of observed covariables is required, and it is necessary to check the order of the estimated linear predictors.

If only some of the covariables are consistent with the assumption of proportional log-odds, the unconstrained partial proportional odds models can be used [15]. In this model only a sub-vector of $\boldsymbol{\beta}$ depends on the categories of $\mathbf{Y}$.

This random effect proportional odds model was already used in various applications: meta-analysis [16], intervention randomized trials in clustered individuals [17] and longitudinal observational studies [18].

The continuation ratio model

When the cumulative probabilities of being in one of the categories greater or equal to *c* in the proportional odds model is replaced by the probability of being in category *c*, this leads to the (backward) continuation ratio model [19]. The response, Y, represents a disease status given by ordered categories with higher values corresponding to more severe disease states. This model can be written as follows:

$$\log\left(\frac{P(Y_{ij}=c \mid \mathbf{X_{ij}}, u_i)}{P(Y_{ij}<c \mid \mathbf{X_{ij}}, u_i)}\right) = \alpha_c + \mathbf{X_{ij}^{'}}\boldsymbol{\beta} + u_i$$

This model is recommended when the underlying outcome is irreversible, in the sense that upon attaining level *c* a subject's response cannot revert to a lower level. In contrast to the proportional odds model, the continuation ratio model does not have the useful property of invariance under reversal or collapse of categories.

The forward continuation ratio model can also be defined by the probability of being in category *c* conditionally of being in category *c* or greater. It is then analogous to the discrete-time version of the proportional hazard model.

As with the proportional odds model several authors introduced more flexibility defining, for each covariable, regression parameters specific to the response category [20, 21].

The generalization with a random effect was used by Dos Santos et al [22] to investigate in a sample of women with breast cancer, the effect of stage (size of the tumor), age and primary treatment on the condition subsequent to surgery (death, progression, disease static, partial response, complete response, no disease) evaluated repeatedly, at one or more follow-ups, until death.

The adjacent category model

This model considers the ratio of the probability of being in one of the categories and the probability of being in the adjacent category. The form of the model is:

$$\log\left(\frac{P(Y_{ij}=c\,|\,\mathbf{X_{ij}},u_i)}{P(Y_{ij}=c-1\,|\,\mathbf{X_{ij}},u_i)}\right)=\alpha_c+\mathbf{X_{ij}'}\boldsymbol{\beta}+u_i$$

In this form the regression parameters do not depend on the level of the response. But like for the models described above, the regression coefficients can vary with *c*, meaning that for a given covariable there would be (*K-1*) parameters $\boldsymbol{\beta_c}$.

This model is designed for situations in which the subject must 'pass through' one category to reach the next.

The stereotype mixed model

In cases in which the ordinal categories of Y cannot be considered as a discrete version of a continuous variable, an extension of the nominal logistic model [23] also called baseline-category model that accounts for the ordering of the response categories can be used. This model is called the stereotype model [24].

$$\log\left(\frac{P(Y_{ij}=c\,|\,\mathbf{X}_{ij},u_i)}{P(Y_{ij}=0\,|\,\mathbf{X}_{ij},u_i)}\right)=\alpha_c+\phi_c\mathbf{X_{ij}'}\boldsymbol{\beta}+u_i \qquad (3)$$

$$with\ 0<\phi_1<\phi_2<\ldots<\phi_k=1$$

An important distinction between the stereotype model and the models described above is that the stereotype model uses comparisons to a reference category.

The constraints on the ϕs completely ensure the response level ordering. The random effects (u_i) and the fixed effects ($\boldsymbol{\beta}$) are then independent of the response level. It is assumed that a single linear combination of covariables is adequate to distinguish between all outcome levels. The distance between the outcome levels in terms of linear predictor is given by the ϕs parameters. More complex stereotype models have been recently introduced using between 2 and (K-2) linear combinations [24]. If two linear combinations are used the model is written:

$$\log\left(\frac{P(Y_{ij}=c\mid \mathbf{X}_{ij},u_i)}{P(Y_{ij}=0\mid \mathbf{X}_{ij},u_i)}\right)=\alpha_c+\phi_c^1\mathbf{X}_{ij}'\boldsymbol{\beta}^1+\phi_c^2\mathbf{X}_{ij}'\boldsymbol{\beta}^2+u_i$$

The nominal logistic model is equivalent to the most flexible stereotype model with (K-1) linear combinations. The stereotype models with less than (K-1) linear combinations may then be compared to the nominal logistic model to see if their simplifying constraints are valid.

Model selection and goodness of fit

Model selection includes the choice of the type of model and variable selection within a model type. In this framework, the parameters estimating method with numerical integration has the advantage of being based on likelihood statistics. Thus, models can be ordered according to likelihood-based measures, such as Akaike's information criterion or Schwarz's Bayesian criterion. Moreover, for variable selection, nested models can be compared with the likelihood ratio test.

However, as in non-longitudinal cases, a more general way of evaluating goodness-of-fit is to calculate the degree to which predicted values agree with observed values. In the generalized linear mixed model, predicted values are calculated from parameter estimates $\left(\hat{\beta}\right)$ and empirical Bayes estimates of random effects ($\hat{u}_i$) [25]. These predicted values can be obtained directly by means of the PREDICT instruction of the SAS

NLMIXED procedure. The concordance between the observed (y_{ij}) and estimated $(\hat{y}_{ij})$ ordinal responses can be assessed by one of the standard measures of ordinal association: gamma, Kendall's tau-b, for example [26]. However, Vonesh suggested that the coefficient of determination and the concordance correlation coefficient could be used [27]. The first one of these coefficients, defined as

$$R^2 = 1 - \sum_i \sum_j \left(y_{ij} - \hat{y}_{ij}\right)^2 / \sum_i \sum_j \left(y_{ij} - \bar{y}\right)^2$$

can be interpreted as the proportion of total variation of **Y** that is accounted for by the fitted model ($\bar{y}$ being the total mean over the subjects and the repetitions of the y_{ij}). The second coefficient is expressed as follows:

$$r_c = 1 - \frac{\sum_i \sum_j \left(y_{ij} - \hat{y}_{ij}\right)^2}{\sum_i \sum_j \left(y_{ij} - \bar{y}\right)^2 + \sum_i \sum_j \left(\hat{y}_{ij} - \bar{\hat{y}}\right)^2 + N\left(\bar{y} - \bar{\hat{y}}\right)^2}$$

This coefficient measures how close the points on a scatter plot of y_{ij} versus $\hat{y}_{ij}$ are to the identity line. Its possible values extend from -1 to 1, with a perfect fit corresponding to a value of one and a lack of fit corresponding to values ≤ 0.

The main advantage is that these criteria are independent of the estimating procedure, making it possible to compare non-nested models or models with different covariance structure or models estimated by the means of non-likelihood based methods.

Application to the EPIDOS study

The French EPIDOS study is a prospective multicenter study of risk factors for hip fractures in 7575 women aged 75 years or older, recruited in 1992-1993 by mailing, based on large population-based listings including electoral rolls [28]. These women were asked to complete an annual postal questionnaire, investigating hospitalization, new health events, changes in weight, type of housing, ability to go outside, activities of daily life (ADL), instrumental activities of daily life (IADL), medication used, and

subjective health evaluation. The present analysis was carried out on the data from Montpellier (Southern France), one of the five participating centers, where 1548 women were followed for 7 years.

Some annual questionnaires were missing, mainly due to illness or to family events, such as the death of the spouse. These women sent in their responses late.

We investigated the risk factors related to changes in IADL score. This score is designed to evaluate independent living and deals with the following activities: performing light or heavy housework, laundry, shopping for groceries or personal items, preparing meals, using a telephone, taking public transport, managing money and taking medication without assistance. This indicator of disability varies from zero to 8 according to the number of activities that the woman is not capable of performing on her own. To ensure that each category contained sufficient numbers, we collapsed the responses as follows: 0, 1-2, 3-4, 5-6, 7-8. The final annual response therefore had five ordered categories and varied from 0 (not disabled) to 4 (almost entirely dependent).

The analysis was restricted to the women free of disability at inclusion and without missing values for the baseline covariables (929 women). We examined only several didactic risk factors to illustrate the use of the mixed ordinal logistic models. We do not present an exhaustive analysis of all covariables potentially involved in the disability process.

Moreover, since the disability process is reversible and may skip a value of the IADL score, the continuation ratio and adjacent category models were not appropriate for the Epidos data. Thus, these models were not used in the analysis.

Model fitting with SAS NLMIXED

The SAS NLMIXED procedure is very flexible, enabling the user to code the likelihood directly. This makes it possible to program all the models described above, as described in this section.

As the response reflects an underlying continuous phenomenon - the decline in the women's general state of health - we used the proportional odds model. As already explained, we began with the generalized ordered logit model (2). In the case of only two covariables (age and time) the SAS code is:

```
1     title1 ' Generalized ordered logit model' ;

2  title2 ' TIME, AGE' ;

3  proc nlmixed data=table qmax=5000  ;
4     bounds i1>0, i2>0, i3>0 ;
5     parms beta11=0.56 beta12=0.61 beta13=0.64 beta14=0.64
6     beta0=-10 beta24=0.3 beta23=0.3 beta22=0.3 beta21=0.3 ;

7     eta4 = beta0 + beta14*time + beta24*age+ u ;
8     eta3 = beta0 + i1 + beta13*time + beta23*age + u  ;
9     eta2 = beta0 + i1 + i2 + beta12*time + beta22*age + u ;
10    eta1 = beta0 + i1 + i2 + i3 + beta11*time + beta21*age + u  ;

11    if (iadl=4) then z=1/(1+exp(-eta4)) ;
12    else if (iadl=3) then z= 1/(1+exp(-eta3)) - 1/(1+exp(-eta4)) ;
13    else if (iadl=2) then z= 1/(1+exp(-eta2)) - 1/(1+exp(-eta3)) ;
14    else if (iadl=1) then z= 1/(1+exp(-eta1)) - 1/(1+exp(-eta2)) ;
15    else if (iadl=0) then z= 1 - 1/(1+exp(-eta1));

16    ll=log(z) ;
17    model iadl ~ general(ll);
18    random u ~ normal(0,su**2) subject=ident;

19    estimate 'int4' beta0 ;
20    estimate 'int3' beta0+i1 ;
21    estimate 'int2' beta0+i1+i2 ;
22    estimate 'int1' beta0+i1+i2+i3 ;

23    predict eta1 out=eta1;
24    predict eta2 out=eta2;
25    predict eta3 out=eta3;
26    predict eta4 out=eta4;
27    run;
```

In this code, "iadl" is the outcome variable (with values from 0 to 4) and "ident" is the variable that indicates the subject ID number (line 18). Lines 7 to 18 define the model

according to equation (2). If there are more than two variables, eta1 to eta4 should be modified and more parameters, beta(i, j), should be used.

We included time since baseline as a continuous variable. We also included the following fixed effects, recorded at inclusion: age at entry (in years exceeding 74), gait speed (in m/s, for the completion of 6 m at a normal pace) and pain (frequent pain in the lower back, hip, knee, ankle or leg). All these covariables except "pain" were considered to be quantitative.

The NLMIXED procedure converged and provided estimations only when the initial parameters, given in the PARMS instruction (line 5), were close to the final solution. We then calculated successive models, beginning with a simple model that includes only a random intercept, introducing the covariables one by one and using the previously estimated parameters as the starting point for the next estimation at each step.

Results

Model selection

The parameters of the full generalized ordered logit model are provided in the first part of Table 1. We reduced the number of parameters by comparing this model with successive restricted models assuming proportional odds ratios for some covariables. We considered each covariable separately: if the confidence interval of one parameter included the parameter of the adjacent category we estimated the model in which the two parameters had been replaced by a single one. This new restricted model was chosen if it did not differ significantly from the initial full model according to the likelihood ratio. For instance, the four parameters of the variable "gait speed" can be replaced by a single parameter (see Table 1). In this case, lines 7 to 10 of the SAS code,

which define the linear predictors, are modified and a single parameter is used to replace the corresponding betas (beta(i, 1), beta(i, 2), beta(i, 3), beta(i, 4)). The final comparisons are shown in Table 2. Comparing the partial proportional odds model with the generalized ordered logit model, the p-value associated with the likelihood ratio was 0.06 (chi2=16.5 with 9 df), indicating that the two models have an equivalent goodness-of-fit. For the covariables time since inclusion and pain, proportionality could be validated for some of the response levels only. For instance the log-odds ratio associated with pain was 0.78 for the first two levels and 0.34 for the next two levels. Pain was found to have a pronounced effect at the beginning of the disability process which became non significant for transitions to more serious states. Time since inclusion was grouped as follows: 4, 3+2, 1. Another solution, equivalent in goodness-of-fit performance, would be: 4+3, 2, 1. But level 2 cannot be grouped with level 4. The effect of time seems to be more marked on the middle categories of disability (2, 3). In contrast, for the covariables "gait speed", and "age", the model was reduced to only one parameter, indicating that the odds are entirely proportional.

The hierarchy of linear predictor estimates was checked for all subjects and all time points. The linear predictors are given by the PREDICT instruction in the SAS program and we checked if the inequality eta1 > eta2 > eta3 > eta4 was verified for all the observations. None presented a reversed order, and the final model was therefore considered valid in the range of the observed values for the covariables.

The generalized ordered logit model and the partial proportional odds model had very similar coefficients of determination (0.58 and 0.57 respectively) and concordance correlation coefficients (0.77 and 0.765). We therefore retained the partial proportional logit model, in accordance with the principle of parsimony.

The goodness-of-fit statistics for the strict proportional odds model are also given in table 2. In comparison with the generalized ordered logit model and the partial proportional odds model, the p values associated with the likelihood ratio were $p<0.0001$ indicating that the proportional odds model was less efficient than the two others. The hypothesis of proportionality for all the covariables was then rejected.

We also fitted a random-effect stereotype model (3) including the same covariables (SAS code given in the appendix). The parameter estimates are given in the last column of Table 1. The reference category was an IADL score equal to zero. In this model each covariable is characterized by only one parameter. The log-odds ratio for a response category in comparison with the reference is obtained by multiplying the estimated parameter by the corresponding estimate of ϕ. For example the log-odds corresponding to a disability level of 3 for the covariable "pain" is equal to 0.705 (0.75*0.94). This model appeared to give the poorest fit: the -2Log Likelihood was 10243, whereas that for the partial proportional odds model was 9790, the coefficient of determination was 0.32 and the concordance correlation coefficient was 0.57 (Table 2). In our example, this model is also less efficient (Table 2) than the proportional odds model.

Goodness of fit

The concordance correlation coefficients reported above were calculated taking into account both the estimates of the fixed parameters and the empirical Bayes estimates of the random effects. They can be seen as conditional model concordance correlation coefficients [27]. The average model concordance correlation coefficient can also be calculated, based entirely on the non-random part of the model. For the final partial proportional odds model, this average model concordance correlation coefficient was 0.37. This coefficient differed from the corresponding conditional model concordance correlation coefficient (equal to 0.77), indicating that the fixed covariables explained

only a part of the response variation. A high level of between-individual heterogeneity is modeled by the random intercept. The probability of being at a given level of disability is therefore partially due to the uncharacterized "frailty" of the woman concerned.

Discussion

The mixed models presented in this paper are interesting in two ways.

Firstly, they allow to analyze ordinal response and to evaluate the influence of each risk factor according to the levels of the response. Compared to the methods that dichotomize the response using an arbitrary cut-off and model it with a simple logistic regression our models are more worthwhile in the sense that the optimal cut-off may change with the risk factors.

Secondly, the random-effect models for non-gaussian responses are very powerful tools for the explanation of longitudinal processes. Several statistical packages (STATA, SAS) have recently implemented algorithms for these methods. The numerical procedure, which directly approximates the likelihood, does not suffer the drawbacks of previously developed methods. For example, the penalized quasi-likelihood approach [29, 30] has been shown to provide highly biased estimates of mixed-effects parameters with binary responses especially when the number of observations by cluster is small [31-33]. But, it is difficult to introduce a complex covariance structure as the computational time greatly increases with the number of random effects. This limitation is likely to become less severe in the future as computer capacity is continually increasing. In our example, the partial proportional logit model presented in Table 1 took 56 minutes on a P4, 1.4GHz PC, and only 12 minutes on a P4, 2.5GHz PC. However the NLMIXED procedure may fail to integrate the likelihood or give non-

stationary estimations. This non-optimal estimation can be diagnosed by checking the gradient (vector of first derivative) of the negative log likelihood function for each parameter. These gradients are systematically provided by SAS in the results. If any one of them is not close to zero, then the solution cannot be considered to be valid. These problems of convergence were not encountered with our set of data.

Other more flexible methods allowing multiple levels of clustering, such as Markov chain Monte Carlo methods [34], can be used but are not based on likelihood statistics and model comparison is less straightforward. However, they can be used to cross-validate the estimations generated by procedures with numerical integration, such as NLMIXED.

Little work has been done on the model checking and the model diagnosis for GLMMs. This is nevertheless an important area in practice and we have presented here a simple method for the comparison of ordinal random-effect models and assessment of their goodness of fit.

Efforts are still required to improve the numerical approximations of likelihood in random-effect modeling, but these methods appear promising.

Conclusions

The models presented in this paper permit the analysis of longitudinal studies that are of particular interest in epidemiology taking into account the ordinal nature of the outcome. Furthermore the impact of the risk factors can be modeled according to the response levels. This presents an important perspective in better explaining complex processes of evolution.

References

1. Katz S., Downs T.D., Cash H.R. Grotz R.C. Progress in development of the index of ADL. Gerontologist 1970; **10**: 20-30.
2. Lawton M.P. Brody E.M. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969; **9**: 179-86.
3. Long J.S. Pavalko E. Comparing alternative measures of functional limitation. Med Care 2004; **42**: 19-27.
4. Simonsick E.M., Kasper J.D., Guralnik J.M., Bandeen-Roche K., Ferrucci L., Hirsch R., *et al.* Severity of upper and lower extremity functional limitation: scale development and validation with self-report and performance-based measures of physical function. WHAS Research Group. Women's Health and Aging Study. J Gerontol B Psychol Sci Soc Sci 2001; **56**: S10-9.
5. Beckett L., Brock D., Lemke J., Mendes De Leon C., Guralnik J., Fillenbaum G., *et al.* Analysis of change in self-reported physical function among older persons in four population studies. Am J Epidemiol 1996; **143**: 766-78.
6. Diggle P., Liang K. Zeger S. Analysis of longitudinal data, Oxford, Clarendon Press: Oxford Science Publications, 1994.
7. Lindsey J. A review of some extensions to generalized linear models. Stat Med 1999; **18**: 2223-36.
8. Zeger S., Liang K. Albert P. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988; **44**: 1049-60.
9. Pinheiro J. Bates D. Approximations to the log-likelihood function in the nonlinear mixed-effects model. J Comput Graph Stat 1995; **4**: 12-35.
10. Ananth C. Kleinbaum D. Regression models for ordinal responses: a review of methods and applications. Int J Epidemiol 1997; **26**: 1323-33.

11. Lall R., Campbell M., Walters S. Morgan K. A review of ordinal regression models applied on health-related quality of life assessments. Stat Methods Med Res 2002; **11**: 49-67.

12. Mc Cullagh P. Regression models for ordinal data (with discussion). J R Statist Soc B 1980; **42**: 109-42.

13. Mc Cullagh P. Nelder J. Generalized Linear Models - Second Edition, Boca Raton, London, New-York, Washington, D.C.: Chapman & Hall/CRC, 1989.

14. Hedeker D. Mermelstein R. A multilevel thresholds of change model for analysis of stages of change data. Multivariate Behavioral Research 1998; **33**: 427-55.

15. Peterson B. Harrell F. Partial proportional odds models for ordinal response variables. Applied Statistics 1990; **39**: 205-17.

16. Whitehead A., Omar R.Z., Higgins J.P., Savaluny E., Turner R.M. Thompson S.G. Meta-analysis of ordinal outcomes using individual patient data. Stat Med 2001; **20**: 2243-60.

17. Marinacci C., Schifano P., Borgia P. Perucci C.A. Application of random effect ordinal regression model for outcome evaluation of two randomized controlled trials. Stat Med 2001; **20**: 3769-76.

18. Leon A.C. Hedeker D. A mixed-effects quintile-stratified propensity adjustment for effectiveness analyses of ordered categorical doses. Stat Med 2005; **24**: 647-58.

19. Armstrong B. Sloan M. Ordinal regression models for epidemiologic data. Am J Epidemiol 1989; **129**: 191-204.

20. Cole S. Ananth C. Regression models for unconstrained, partially or fully constrained continuation odds ratios. Int J Epidemiol 2001; **30**: 1379-82.

21. Greenland S. Alternative models for ordinal logistic regression. Stat Med 1994; **13**: 1665-77.

22. Dos Santos D.M. Berridge D.M. A continuation ratio random effects model for repeated ordinal responses. Stat Med 2000; **19**: 3377-88.

23. Engel J. Polytomous logistic regression. Stat Neerlandica 1988; **42**: 233-52.

24. Lunt M. Prediction of ordinal outcomes when the association between predictors and outcome differs between outcome levels. Stat Med 2005; **24**: 1357-69.

25. Carlin B. Louis T. Bayes and empirical Bayes methods for data analysis- Second Edition, ed. H.C. Chapman, London, 2000.

26. Agresti A. Categorical data analysis, New York, Chichester, Brisbane, Toronto, Singapore: Wiley -Interscience Publication, 1990.

27. Vonesh E., Chinchilli V. Pu K. Goodness-of-fit in generalized nonlinear mixed-effects models. Biometrics 1996; **52**: 572-87.

28. Dargent-molina P., Favier F., Grandjean H., Baudoin C., Schott A., Hausherr E., *et al.* Fall-related factors and risk of hip fracture: the EPIDOS prospective study. Lancet 1996; **348**: 145-49.

29. Breslow N. Clayton D. Approximate inference in generalized linear mixed models. J Am Stat Assoc 1993; **88**: 9-25.

30. Wolfinger R. O'connell M. Generalized linear mixed models: a pseudo-likelihood approach. J Statist Comput Simul 1993; **48**: 233-43.

31. Breslow N. Bias correction in generalised linear mixed models with a single component of dispersion. Biometrika 1995; **82**: 81-91.

32. Rodriguez G. Goldman N. An assessment of estimation procedures for multilevel models with binary responses. J R Statist Soc A 1995; **158**: 73-89.

33. Heo M. Leon A.C. Comparison of statistical methods for analysis of clustered binary observations. Stat Med 2005; **24**: 911-23.

34. Zeger S. Karim M. Generalized linear models with random effects: A Gibbs sampling approach. J Am Stat Assoc 1991; **86**: 79-86.

Acknowledgements

We would like to thank Dr François Favier and the EPIDOS group who provided the data for this work.

Table I. - Parameter estimates of three random-effect models for ordinal responses

	Generalized ordered logit model			Partial proportional odds model			Stereotype model		
	Parameter	SE	p value	Parameter	SE	p value	Parameter	SE	p value
Intercept									
4	-5.19	0.74	< 0.0001	-5.99	0.56	< 0.0001	-4.05	0.55	< 0.0001
3	-4.82	0.63	< 0.0001	-5.00	0.52	< 0.0001	-3.76	0.53	< 0.0001
2	-3.66	0.56	< 0.0001	-3.58	0.51	< 0.0001	-2.62	0.46	< 0.0001
1	-0.73	0.52	0.16	-0.63	0.51	0.22	-1.21	0.31	< 0.0001
Time since inclusion (years)									
4	0.57	0.05	< 0.0001	0.59	0.05	< 0.0001	0.79	0.04	< 0.0001
3	0.61	0.04	< 0.0001	0.64	0.03	< 0.0001			
2	0.65	0.03	< 0.0001	"	"	"			
1	0.52	0.02	< 0.0001	0.53	0.02	< 0.0001			
Age (in years minus 74)									
4	0.25	0.04	< 0.0001	0.24	0.03	< 0.0001	0.35	0.03	< 0.0001
3	0.27	0.03	< 0.0001	"	"	"			
2	0.27	0.03	< 0.0001	"	"	"			
1	0.22	0.03	< 0.0001	"	"	"			
Gait speed (m/s)									
4	-4.73	0.73	< 0.0001	-3.94	0.50	< 0.0001	-5.26	0.58	< 0.0001
3	-4.22	0.61	< 0.0001	"	"	"			
2	-4.14	0.55	< 0.0001	"	"	"			
1	-3.73	0.51	< 0.0001	"	"	"			
Pain									
4	0.04	0.25	0.88	0.34	0.21	0.10	0.75	0.19	< 0.0001
3	0.31	0.22	0.15	"	"	"			
2	0.73	0.19	0.0002	0.78	0.17	< 0.0001			
1	0.83	0.18	< 0.0001	"	"	"			
Scale parameters									
ϕ_4							1		
ϕ_3							0.94	0.05	< 0.0001
ϕ_2							0.83	0.04	< 0.0001
ϕ_1							0.55	0.03	< 0.0001
Random effect SD	2.28	0.08	<0.0001*	2.28	0.08	< 0.0001*	2.28	0.10	< 0.0001*

* Stram and Lee test [35]

Table II - Goodness of fit of four random-effect models for ordinal responses

	Generalized ordered logit model	Partial proportional odds model	Proportional odds model	Stereotype model
-2LL	9773	9790	9818	10243
Number of parameters	21	12	9	12
SBC	9916	9872	9880	10325
Coefficient of determination	0.580	0.571	0.579	0.316
Concordance correlation coefficient	0.770	0.765	0.768	0.570

Appendix :

SAS code for the stereotype model:

```
title1 ' Stereotype' ;
title2 ' AGE, TIME' ;

proc nlmixed data= table qmax=5000  ;

      bounds 0 < epsy1-epsy3 <= 0.99  ;

      parms alpha4=-8 alpha3=-8 alpha2=-6 alpha1=-4
      beta1=0.8 beta2=0.4
      epsy3=0.1 epsy2=0.1 epsy1=0.2  ;

      eta4 = alpha4 + beta1*time + beta2*age + u  ;
      eta3 = alpha3 + (1-epsy3)*(beta1*time + beta2*age) + u  ;
      eta2 = alpha2 + (1-epsy3-epsy2)*(beta1*time + beta2*age) + u ;
      eta1 = alpha1 + (1-epsy3-epsy2-epsy1)*(beta1*time + beta2*age) +
u  ;

      if (iadl=4) then
      z=exp(eta4)/(exp(eta1)+exp(eta2)+exp(eta3)+exp(eta4)+1) ;
      else if (iadl=3) then z=exp(eta3)/(exp(eta1)+exp(eta2)+
      exp(eta3)+exp(eta4)+1) ;
      else if (iadl=2) then z=exp(eta2)/(exp(eta1)+exp(eta2)+
      exp(eta3)+exp(eta4)+1) ;
      else if (iadl=1) then z=exp(eta1)/(exp(eta1)+exp(eta2)+
      exp(eta3)+exp(eta4)+1) ;
      else if (iadl=0) then z=1/(exp(eta1)+exp(eta2)+
      exp(eta3)+exp(eta4)+1) ;

      ll= log(z) ;
      model iadl ~ general(ll) ;
      random u ~ normal(0,su**2) subject=ident;

      estimate 'phi3' (1-epsy3);
      estimate 'phi2' (1-epsy3-epsy2);
      estimate 'phi1' (1-epsy3-epsy2-epsy1);

      predict eta1 out=eta1;
      predict eta2 out=eta2;
      predict eta3 out=eta3;
      predict eta4 out=eta4;

run ;
```