STATISTICS IN MEDICINE, VOL. 8, 1209–1224 (1989)

A SURVEY OF MODELS FOR REPEATED ORDERED CATEGORICAL RESPONSE DATA

ALAN AGRESTI

Department of Statistics, University of Florida, Gainesville, FL 32611, U.S.A.

SUMMARY

We survey models for analysing repeated observations on an ordered categorical response variable. The models presented are univariate models that permit correlation among repeated measurements. The models describe simultaneously the dependence of marginal response distributions on values of explanatory variables and on the occasion of response. We present models for three transformations of the response distribution: cumulative logits, adjacent-category logits, and the mean for scores assigned to response categories. We discuss three methods for fitting the models: maximum likelihood, weighted least squares, and semi-parametric. Weighted least squares is easily implemented with SAS, as illustrated with a study designed to compare a drug with a placebo for the treatment of insomnia.

KEY WORDS Cumulative logits Logit and log-linear models Longitudinal data Marginal homogeneity Ordinal data Repeated measures

1. INTRODUCTION

Many studies measure a response variable for each subject repeatedly, resulting in dependent responses. For instance, after patients begin treatment for some disease, a clinician might periodically evaluate their response to the treatment. In health-related applications with repeated measurement, the response variable is often *categorical.* Clinicians might simply evaluate whether a treatment is successful. This article considers *ordered* categorical response variables, such as evaluation of treatment success measured on a four-point scale: excellent, good, fair, poor.

Koch *et al.*[1] wrote one of the first articles to describe types of repeated categorical response data and possible models. Recent articles such as those by Landis *et al.*,[2] Ware, Lipsitz and Speizer[3] and Stram, Wei and Ware[4] have refocused attention on this topic. Our discussion includes models of the sort described in these papers, plus similar ones related to log-linear models. Though the methodology we discuss is not new, we hope that an organized presentation of potential models and methods of fitting them will be helpful to applied statisticians. In addition, we highlight some problems that may stimulate future research.

Table I, taken from Francom, Chuang and Landis,[5] is an example of repeated ordered categorical response data. The table shows results of a randomized, double-blind clinical trial comparing an active hypnotic drug with a placebo in patients with insomnia. The outcome variable is patient response to the question 'How quickly did you fall asleep after going to bed?', using categories (<20, 20–30, 30–60, >60) minutes. Patients responded at the start and conclusion of a two-week treatment period. The repeated measurement makes the response bivariate, measured at levels: initial, follow-up. We refer to these levels as *occasions.* We regard the

0277–6715/89/101209–16$08.00

Received September 1988
Revised April 1989

Table I. Frequency distribution of time to falling asleep (minutes), by treatment and occasion

Treatment	Initial occasion	Follow-up occasion <30	20–30	30–60	>60
Active	<20	7	4	1	0
	20–30	11	5	2	2
	30–60	13	23	3	1
	>60	9	17	13	8
Placebo	<20	7	4	2	1
	20–30	14	5	1	0
	30–60	6	9	18	2
	>60	4	11	14	22

Source: Francom, Chuang and Landis.[5]

treatments, active and placebo, as levels of a binary explanatory variable. The subjects receiving the two different treatments are independent samples.

One approach for repeated categorical data involves analysing patterns of change for individual subjects, by modelling cell probabilities in the joint distribution of the repeated response. Ware, Lipsitz and Speizer[3] referred to this approach as *transitional* modelling. For descriptive and inferential purposes, there is often less interest in the multivariate dependence among repeated responses than in population characteristics of the response.[1,4,6] A second approach involves analysing patterns of change for populations, by modelling marginal distributions. This permits investigating questions such as 'For a given treatment, does the response distribution change across occasions?', or 'At a given occasion, are there differences among response distributions for the treatments?', or 'Is the difference for two treatments the same at all occasions?' Ware, Lipsitz and Speizer[3] reviewed conceptual and technical differences between marginal and transitional approaches. In this article, we focus primarily on the marginal approach, but note instances in which that approach is not fully informative.

We define the *link* to be the transformation of the response distribution that is modelled, and the *linear predictor* to be the linear combination of explanatory variables that the model relates to the link. We borrow this terminology from generalized linear models, though the usage of link here is a multivariate generalization of the usual one.[7] The process of formulating and fitting a model is one whereby we select (a) the link, (b) the linear predictor, and (c) the method for estimating parameters in the resulting model. Sections 2–4 present some choices in this process. Section 2 presents links for ordinal response variables, with emphasis on cumulative logits, adjacent-category logits, and a mean for scored response categories. Section 3 presents a hierarchy of linear predictors for describing occasion and covariate effects. Section 4 discusses three methods for model fitting: maximum likelihood, weighted least squares, and a semi-parametric one that more easily handles sparse or missing data and time-dependent covariates.

Section 5 illustrates some models, using the cumulative logit link to analyse Table I. Sections 6 and 7 present similar analyses using adjacent-category logits and a mean response. Section 8 analyses Table I using transitional models. Section 9 discusses software availability for the models, and illustrates use of SAS to perform analyses reported in Sections 5–7. The final section outlines a more fully multivariate approach, one that describes changes in marginal distributions at the subject level rather than population level, and suggests future research for fitting such models.

2. LINKS FOR ORDINAL RESPONSES

Suppose each subject may be observed at d occasions, and let $(1, 2, \ldots, r)$ denote the r possible response categories at each occasion. The data can be described by a contingency table with r^d cells, containing counts of possible multivariate response profiles. Let

$$\pi_{\mathbf{j}} \qquad \text{with } \mathbf{j}=(j_1, \ldots, j_d)$$

denote the probability that a randomly selected subject makes response j_g at occasion g, $1 \leqslant j_g \leqslant r$, $g=1, \ldots, d$. Let the $+$ subscript denote summation over an index. Then $\{\pi_{+\ldots+k+\ldots+},\ k=1, \ldots, r\}$, where k is in position g, is the marginal distribution of the response at occasion g. Denote these marginal probabilities by $\{\phi_{gk},\ k=1, \ldots, r\}$, with $\Sigma_k \phi_{gk}=1$. Denote the response for a randomly selected subject at occasion g by Y_g, so that

$$\phi_{gk}=P(Y_g=k), \qquad k=1, \ldots, r.$$

We discuss models that apply some selected link to each marginal distribution. The logit link is defined for binary responses, but can be applied with $r>2$ response categories. There are $r-1$ non-redundant logits of any given type. We take order of response categories into account by constructing logits for cumulative probabilities

$$\text{logit}[P(Y_g \leqslant k)]=\log[(\phi_{g1}+\ \ldots\ +\phi_{gk})/(\phi_{g,k+1}+\ \ldots\ +\phi_{gr})], \qquad k=1, \ldots, r-1,$$

called *cumulative logits*, and for adjacent-response probabilities

$$\text{logit}[P(Y_g=k \mid Y_g=k \text{ or } k+1)]=\log[\phi_{gk}/\phi_{g,k+1}], \qquad k=1, \ldots, r-1,$$

called *adjacent-category logits*.

For categorical (nominal or ordinal) explanatory variables, models for adjacent-category logits are equivalent to log-linear models with scores assigned to levels of ordinal variables.[8] McCullagh[9] popularized models using cumulative logits. This logit is one member of a family of *cumulative links*, strictly monotone functions that transform the (0, 1) scale for cumulative probabilities onto the real line, the scale for linear predictors. Though we specify models using logits, the models also make sense for other links. For instance, for the probit link one models $\Phi^{-1}[P(Y_g \leqslant k)]$ or $\Phi^{-1}[P(Y_g=k \mid Y_g=k \text{ or } k+1)]$, where Φ is the standard normal CDF.

When there are covariates $\mathbf{x}$, we allow a separate distribution

$$\{\pi_{\mathbf{j}}(\mathbf{x}),\ 1 \leqslant j_g \leqslant r,\ g=1, \ldots, d\}$$

at each level of $\mathbf{x}$. For Table I, for instance, $d=2$, $r=4$, $\mathbf{x}$ is the binary classification (active, placebo), and the table consists of two r^d $(=4 \times 4)$ components. We then denote the marginal distributions by $\{\phi_{gk}(\mathbf{x}),\ k=1, \ldots, r\}$, and apply the link to them at each $\mathbf{x}$.

The next section formulates models that apply simultaneously to $r-1$ transformations of response probabilities at each occasion. For the cumulative logit link, we motivate the models by appealing (as in Anderson and Philips[10]) to a regression model for an underlying continuous response. Let Y^* denote an underlying continuous variable having CDF $G(y-\eta)$, where η is a location parameter dependent on the occasion and covariate through $\eta(\mathbf{x})=\mu_g+\boldsymbol{\beta}'\mathbf{x}$. Suppose we cannot observe Y^*, but $-\infty=\alpha_0<\alpha_1<\ldots<\alpha_r=\infty$ are such that

$$Y=k \qquad \text{if } \alpha_{k-1}<Y^* \leqslant \alpha_k.$$

The $\{\alpha_k\}$ are called *cutpoint* parameters. Then at occasion g

$$P(Y \leqslant k)=P(Y^* \leqslant \alpha_k)=G(\alpha_k-\mu_g-\boldsymbol{\beta}'\mathbf{x}\}.$$

Thus a model for cumulative probabilities holds with link equal to G^{-1}. For instance, if $Y^* = \mu_g + \boldsymbol{\beta}'\mathbf{x} + \varepsilon$, with ε having a logistic distribution, then G^{-1} is the logit transform, and the proper model utilizes cumulative logits.

As a consequence of this construction, in cumulative logit models we use linear predictors in which effects of occasions and explanatory variables are the same for each cutpoint; that is, $\{\mu_g\}$ and $\boldsymbol{\beta}$ do not vary according to k. In models with adjacent-category logit link, we also use effects that are constant by cutpoint. Such behaviour occurs whenever there is an underlying normal response with constant variance and the categories are equally spaced. This follows from Goodman,[11] who showed that a related model for two-way tables is a discrete analogue of the bivariate normal distribution.

3. A HIERARCHY OF LINEAR PREDICTORS

We now give linear predictors that describe simultaneously how the marginal distribution changes across occasions $g = 1, \ldots, d$ for fixed $\mathbf{x}$, and how it depends on $\mathbf{x}$ for fixed g. The linear predictors are hierarchical, describing occasion and covariate effects with varying degrees of generality. Interpretations depend on whether the difference between marginal distributions for any two covariate values is the same for all occasions; that is, on whether there is a lack of 'occasion × covariate interaction'. For a given model, we define the residual degrees of freedom (d.f.) as the difference between the number of response functions (for example, the total number of cumulative logits at the various combinations of g, k, and $\mathbf{x}$) and the number of parameters in the model.

We begin with the simple case of no covariates. That is, we model the behaviour over d occasions of a single population. The marginal distributions of the contingency table form a $d \times r$ table of $\{\phi_{gk}\}$ values. Let L_{gk} denote the linear predictor for the link evaluated at cutpoint k of the marginal distribution at occasion g. The model of marginal homogeneity is

$$L_{gk} = \alpha_k, \qquad g = 1, \ldots, d, \; k = 1, \ldots, r-1. \tag{1}$$

The saturated model for the marginal distributions, $L_{gk} = \alpha_{gk}$, permits marginal heterogeneity.

It is sensible to model potential heterogeneity, such as by

$$L_{gk} = \alpha_k + \mu_g. \tag{2}$$

For identifiability, we impose a constraint on the $\{\mu_g\}$, such as $\mu_d = 0$. The cutpoints $\{\alpha_k\}$ are nuisance parameters, and the occasion effects are described by $\{\mu_g\}$. There are $d(r-1)$ marginal link values, and model (2) has $(r-1)+(d-1)$ parameters, so the residual d.f. $= (d-1)(r-2)$. Model (1) of marginal homogeneity has residual d.f. $= (d-1)(r-1)$.

From motivation given in the previous section, model (2) assumes effects are identical for all cutpoints k. For instance,

$$L_{bk} - L_{ak} = \mu_b - \mu_a, \qquad k = 1, \ldots, r-1.$$

For the cumulative logit link, this means the odds that a randomly selected subject at occasion b makes response $\leqslant k$ are $\exp(\mu_b - \mu_a)$ times greater than the corresponding odds for a randomly selected subject at occasion a.

Generalizations of model (2) describe occasion effects and covariate effects simultaneously, for instance through the model

$$L_{gk}(\mathbf{x}) = \alpha_k + \mu_g + \boldsymbol{\beta}'\mathbf{x}. \tag{3}$$

Table II. Summary of models with homogeneous cutpoint effects

Model $L_{gk}(\mathbf{x})=$	Description of model	Residual d.f.
1. $\alpha_k+\mu_g+\boldsymbol{\beta}_g'\mathbf{x}$	occasion × covariate interaction	$(r-1)\,(ds-1)-(d-1)-\Sigma_g\dim(\boldsymbol{\beta}_g)$
2. $\alpha_k+\mu_g+\boldsymbol{\beta}'\mathbf{x}$	covariate and occasion effects	$(r-1)\,(ds-1)-(d-1)-\dim(\boldsymbol{\beta})$
3. $\alpha_k+\boldsymbol{\beta}'\mathbf{x}$	covariate effects, no occasion effects	$(r-1)\,(ds-1)-\dim(\boldsymbol{\beta})$
4. $\alpha_k+\mu_g$	occasion effects, no covariate effects	$(r-1)\,(ds-1)-(d-1)$
5. α_k	no occasion effects, no covariate effects	$(r-1)\,(ds-1)$

For links of cumulative probabilities, the property

$$L_{bk}(\mathbf{x}_1)-L_{ak}(\mathbf{x}_2)=(\mu_b-\mu_a)+\boldsymbol{\beta}'(\mathbf{x}_1-\mathbf{x}_2) \tag{4}$$

for model (3) induces an ordering of those probabilities among levels of $\mathbf{x}$ and occasions. Since the ordering is the same for each k, there is a stochastic ordering of the marginal distributions at all occasion–covariate combinations.

A generalization of model (3) replaces $\boldsymbol{\beta}$ by $\boldsymbol{\beta}_g$. This permits occasion × covariate interaction but maintains a simple structure for the marginal heterogeneity. For fixed $\mathbf{x}$, the marginal distributions are location shifts on the scale of the link. Table II lists a variety of related linear predictors that have occasion and covariate effects independent of the cutpoint.

The explanatory variables $\mathbf{x}$ in models such as (3) can be continuous or discrete, though some statistical computer packages can handle only the latter case. Suppose $\mathbf{x}$ is fully categorical, and denote by s the number of settings of $\mathbf{x}$ at which observations occur. Then the data consist of cell counts in a $s\times r^d$ contingency table. For this case, Table II lists residual degrees of freedom for testing model goodness of fit, where $\dim(\boldsymbol{\beta}_g)$ and $\dim(\boldsymbol{\beta})$ denote the dimension of the component of the model parameter vector that describes covariate effects. These formulas are based on $ds(r-1)$ values of marginal links and an identifiability constraint for $\{\mu_g\}$.

4. ESTIMATION AND MODEL FITTING

Suppose there is multinomial sampling over the r^d possible response profiles, with independent samples at each of the s levels of $\mathbf{x}$. In other words, if $\boldsymbol{\pi}$ denotes the cell probabilities, then $\boldsymbol{\pi}$ consists of s independent sets of multinomial probabilities. One can use weighted least squares (WLS), maximum likelihood (ML), or semi-parametric methods to fit models.

Haber[12, 13] gave iterative Newton–Raphson routines for obtaining ML estimates of parameters in models of the form

$$\mathbf{A}\boldsymbol{\pi}=\mathbf{X}\boldsymbol{\beta}$$

or

$$\mathbf{A}\log\mathbf{B}\boldsymbol{\pi}=\mathbf{X}\boldsymbol{\beta}. \tag{5}$$

Mean response models presented in Section 7 have the first form. Models with cumulative logit or adjacent-category logit link have the second form. For instance, for adjacent-category logits, $\mathbf{B}$

contains '0' and '1' elements such that $\mathbf{B}\boldsymbol{\pi}$ produces the rd marginal probabilities (that is, r response probabilities at each of the d occasions) for each level of $\mathbf{x}$. For cumulative logits, $\mathbf{B}$ produces the $2(r-1)d$ cumulative marginal probabilities and their complements. In either case, the log transform is applied to all elements in $\mathbf{B}\boldsymbol{\pi}$, and each row of the matrix $\mathbf{A}$ contains '0' elements except for a single '1' and '-1' positioned to form a particular logit.

Haber's routines apply Aitchison and Silvey[14] methods for maximizing a likelihood subject to constraints. For the ML fit, standard likelihood methods apply for making inferences. For example, marginal homogeneity model (1) is model (2) with $\mu_1 = \ldots = \mu_d$. Twice the difference in maximized likelihoods for the two models is a statistic for testing marginal homogeneity. Haber's routines are impractical when the table has a large number of cells (which happens when there are several occasions, several levels of $\mathbf{x}$, and a multipoint scale), because of the size of the matrix that requires inversion.

Koch *et al.*[1] presented a WLS approach for repeated measures data, which we now outline. Suppose $\mathbf{p}$ is the sample proportion estimate of $\boldsymbol{\pi}$, and let $\mathbf{V}$ denote the sample covariance matrix of $\mathbf{p}$. When $\mathbf{x}$ has s levels, $\mathbf{V}$ is an s-block diagonal matrix with separate multinomial covariance structure for each block. By the delta method the sample responses ($\mathbf{A}$ log $\mathbf{Bp}$) for model (5) have approximate covariance $\mathbf{S} = \mathbf{AD}^{-1}\mathbf{BVB}'\mathbf{D}^{-1}\mathbf{A}'$, where $\mathbf{D}$ is the diagonal matrix with $\mathbf{Bp}$ on the main diagonal. The WLS estimate of $\boldsymbol{\beta}$ is $\mathbf{b} = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}$ ($\mathbf{A}$ log $\mathbf{Bp}$), and $(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}$ estimates the asymptotic covariance matrix of $\mathbf{b}$.

The WLS approach tests goodness of fit using the quadratic form

$$(\mathbf{A} \log \mathbf{Bp} - \mathbf{Xb})'\mathbf{S}^{-1} (\mathbf{A} \log \mathbf{Bp} - \mathbf{Xb}).$$

Wald statistics are used for hypothesis testing. For instance, for testing marginal homogeneity with model (2), the WLS approach can use $\mathbf{c}'[\hat{\mathrm{C}}\mathrm{ov}(\mathbf{c})]^{-1}\mathbf{c}$, where

$$\mathbf{c} = (\hat{\mu}_1 - \hat{\mu}_2, \hat{\mu}_2 - \hat{\mu}_3, \ldots, \hat{\mu}_{r-1} - \hat{\mu}_r)'$$

and $n\hat{\mathrm{C}}\mathrm{ov}(\mathbf{c})$ is the WLS estimator of the covariance of $\sqrt{n}\mathbf{c}$.

A disadvantage of WLS is its inefficiency in handling continuous covariates. These must be collapsed into categorical variables in order to estimate the multinomial sampling structure at each of the s levels of $\mathbf{x}$. In addition, when there are several explanatory variables, s is large, and there may be few observations in each table of r^d possible response patterns. The WLS approach requires a non-singular estimated covariance matrix for the $ds(r-1)$ sample marginal links for the models in Section 3. There are often difficulties with tables having small sample sizes or tables that are large and sparse, because of possible ill-defined sample links or a singular sample covariance matrix for the sample response functions.

WLS methods are more manageable for marginal models than for transitional models for cell probabilities in the full $s \times r^d$ table. Even if the table interior is quite sparse, the $s \times r \times d$ table of marginal counts may not be. Thus, the WLS approach is adequate for marginal models applied to data such as Table I. WLS is an approximation for ML, which corresponds to iteratively reweighted least squares. WLS is asymptotically equivalent to ML, when s is fixed and the sample size grows unboundedly at each level of $\mathbf{x}$.

Complications that often occur for repeated categorical measurement include missing data, time-dependent covariates, and a sampling design more complex than independent multinomial. For WLS analyses, Stanish, Gillings and Koch,[15] Woolson and Clarke,[16] and Lipsitz[17] discussed handling of missing data. For cumulative logit models, Landis *et al.*[18] incorporated sampling weights and design effects into test statistics, using Taylor-series approximations to obtain weighted proportions and their covariance matrix.

ML and WLS statistics for testing marginal homogeneity have asymptotic null chi-squared distributions with d.f. $= d-1$. Standard chi-squared statistics for testing marginal homogeneity (see Darroch[19]) treat classifications as nominal, and have d.f. $= (d-1)(r-1)$. Since the statistics described here are directed at narrower alternatives that reflect the ordering of categories, they are more powerful than the standard statistics when models such as (3) hold. They are reasonable statistics for detecting location differences in marginal distributions, even when such models hold only approximately.

Stram, Wei and Ware[4] proposed an alternative to ML and WLS for cumulative logit modelling of repeated measures data. Their approach is semi-parametric, not assuming a model of dependence among the repeated observations. One fits a cumulative logit model separately to each of the $s \times r$ marginal tables obtained for the d different occasions. One estimates empirically the covariance matrix of the separate estimates of covariate effects, using the approximate linearity in the observations of the effect estimates. The separate estimates are combined in a Wald statistic to test for occasion $\times$ covariate interaction or to estimate a common covariate effect over the occasions.

The focus in Stram, Wei and Ware[4] is on estimation of covariate effects, rather than occasion effects (such as the $\{\mu_g\}$ in model (3)). Their approach yields estimates of cutpoint parameters $\{\alpha_{gk}\}$ for the various occasions, but they are treated as nuisance parameters. Presumably one could use the semi-parametric methodology to obtain estimated covariances for the cutpoint estimates, and hence to fit a structure such as $\alpha_{gk} = \alpha_k + \mu_g$ and test equality of the $\{\mu_g\}$. Also, one could in principle use semi-parametric methodology for alternative links, such as adjacent-category logits. This is not possible, however, with current software.

Another semi-parametric approach can be developed as an extension of the Liang and Zeger[6] methodology for repeated binary responses (see also Chapter 5 of Lipsitz[17]). The model parameters are estimated as if the repeated observations were independent; that is, each subject contributes d independent r-level multinomials to a set of estimating equations. The parameter estimates are consistent and asymptotically normal, but the inverse of the estimated information matrix is not consistent for the true asymptotic covariance matrix. However, one can obtain consistency using a 'robust' estimate such as that proposed by Liang and Zeger.[6] This approach has not yet been developed in the literature.

The ML, WLS, and semi-parametric approaches each have certain advantages. The ML and WLS approaches have the elegance of simultaneously describing occasion and covariate effects. Also, though the semi-parametric approaches make no assumption about dependence structure, they may be less efficient than ML or WLS approaches in estimating effects if the multinomial model truly holds. On the other hand, compared to ML and WLS approaches, the semi-parametric approaches make it simpler to allow for time-dependent covariates and for missing data. Also, they are applicable with large or sparse tables for which use of ML or WLS might be infeasible. If the data are sparse or have many missing components or time-dependent covariates, the semi-parametric approaches are more practical than ML or WLS. For further discussion of the advantages and disadvantages of these approaches, see Ware, Lipsitz and Speizer.[3]

WLS is more accessible than ML or the semi-parametric approaches for fitting marginal models with current statistical computer packages. Thus statisticians may find it to be the currently most practical option for analysing simple tables such as Table I. We used SAS[20] (PROC CATMOD) to perform the WLS analyses of Table I reported in the next three sections.

5. EXAMPLE, USING CUMULATIVE LOGIT LINK

In this section we analyse Table I using models with cumulative logit link fitted by WLS. Table III contains the sample marginal distributions for the four combinations of treatment and occasion.

Table III. Observed and (in parentheses) fitted marginal proportions for cumulative logit model

Treatment	Occasion	Response <20	20–30	30–60	>60
Active	Initial	0·101 (0·102)	0·168 (0·184)	0·336 (0·303)	0·395 (0·411)
	Follow-up	0·336 (0·385)	0·412 (0·303)	0·160 (0·200)	0·092 (0·111)
Placebo	Initial	0·117 (0·098)	0·167 (0·179)	0·292 (0·301)	0·425 (0·421)
	Follow-up	0·258 (0·239)	0·242 (0·286)	0·292 (0·273)	0·208 (0·202)

The treatments have similar distributions at the initial occasion. From the initial to follow-up occasion, the sample distribution of time to falling asleep shifts downwards for both treatments. The degree of shift seems greater for the active drug, though, indicating possible interaction. Since there is only a single covariate (treatment), we replace $L_{gk}(\mathbf{x})$ by the simpler notation L_{gki}, where i indexes treatment.

We allow for interaction with the model

$$L_{gki}=\alpha_k+\mu_g+\beta_i+\eta_{gi}, \qquad g=1, 2,\ k=1, 2, 3,\ i=1, 2, \tag{6}$$

where we constrain $\mu_2=\beta_2=\eta_{12}=\eta_{21}=\eta_{22}=0$. The WLS fit gives $\hat{\mu}_1=-1{\cdot}05$ (asymptotic standard error $=0{\cdot}16$), $\hat{\beta}_1=0{\cdot}69$ (ase $=0{\cdot}23$), and $\hat{\eta}_{11}=-0{\cdot}65$ (ase $=0{\cdot}25$). There is strong evidence of interaction. At the initial occasion, the odds that falling asleep is below any fixed level is estimated to be $\exp(0{\cdot}69-0{\cdot}65)=1{\cdot}04$ times as high for the active drug as for the placebo; at the follow-up, the effect is $\exp(0{\cdot}69)=2{\cdot}0$. At the latter occasion, subjects taking the active drug tended to fall asleep more quickly than those taking placebo.

Model (6) implies marginal homogeneity over occasions for the active treatment when $\mu_1+\eta_{11}=0$, and for the placebo when $\mu_1=0$. There is substantial evidence of heterogeneity for each treatment. For the placebo, for instance, $\hat{\mu}_1/(\text{ase})=6{\cdot}6$. For the placebo treatment, the odds that time to falling asleep is below any fixed level is estimated to be $\exp(1{\cdot}05)=2{\cdot}9$ times as high at the follow-up occasion as at the initial; for the active treatment, the effect is $\exp(1{\cdot}05+0{\cdot}65)=5{\cdot}5$. The model gives a simple and economical description of variation among the four marginal distributions. It has a residual chi-squared statistic of 7·4, with degrees of freedom equal to 6, since there are 12 cumulative logits (3 for each marginal distribution) and 6 parameters ($\alpha_1, \alpha_2, \alpha_3, \mu_1, \beta_1, \eta_{11}$). Table III also contains the fitted marginal distributions. Lack of fit occurs for response 20–30 for the active treatment and follow-up occasion, the dispersion in response times being somewhat less than predicted by the model.

6. ADJACENT-CATEGORY LOGIT MODELS

Goodman[8] used adjacent-category logits in models for a single response. Those models are equivalent to log-linear models for ordinal variables.[11] Adjacent-category logit models for marginal probabilities $\{\phi_{gk}(\mathbf{x})\}$ have similar log-linear model representations. For instance, suppose there is a single covariate. Let ϕ_{gki} denote $\phi_{gk}(\mathbf{x})$ for level i of $\mathbf{x}$. Model (3), applied with

Table IV. Summary of results for logit models fitted to Table I (ase values in parentheses)

Model	Effect	Cumulative logit	Adjacent categories logit
No interaction	Treatment	0·37 (0·20)	0·20 (0·11)
	Occasion	−1·29 (0·13)	−0·67 (0·07)
	Residual χ^2	14·5, d.f. = 7	15·4, d.f. = 7
Interaction	Treatment	0·69 (0·23)	0·38 (0·13)
	Occasion	−1·05 (0·16)	−0·55 (0·09)
	Treat × Cond.	−0·65 (0·25)	−0·36 (0·14)
	Residual χ^2	7·4, d.f. = 6	8·3, d.f. = 6

adjacent-category logits, is

$$\log(\phi_{gki}/\phi_{g,k+1,i}) = \alpha_k + \mu_g + \beta_i. \tag{7}$$

The ds marginal distributions are stochastically ordered according to the values of $\{\mu_g + \beta_i\}$. Model (7) is equivalent to log-linear model

$$\log(\phi_{gik}) = \mu + \lambda_g^{O} + \lambda_i^{X} + \lambda_k^{R} + \lambda_{gi}^{OX} - a_k\beta_i - a_k\mu_g$$

with $a_k = k$ and $\alpha_k = \lambda_k^{R} - \lambda_{k+1}^{R}$, where O = occasion, R = response, and X = covariate. This log-linear model applies to the $d \times s \times r$ marginal probabilities of the original $s \times r^d$ contingency table. Because of the dependence in responses across occasions, it cannot be fitted using standard log-linear methods for three-way tables.

Table IV summarizes WLS fitting of Table I by model (7) and the related interaction model. We interpret results for the latter model as follows. Initially, the odds that time to falling sleep is <20 minutes instead of 20–30 minutes (or 20–30 minutes instead of 30–60 minutes, or 30–60 minutes instead of >60 minutes) is estimated to be $\exp(0{\cdot}38 - 0{\cdot}36) = 1{\cdot}01$ times as high for the active drug group as for the placebo group; at follow-up, the effect is $\exp(0{\cdot}38) = 1{\cdot}46$. For the active drug, the odds that time to falling asleep is in category k instead of $k+1$ ($k = 1, 2, 3$) is estimated to be $\exp(0{\cdot}55 + 0{\cdot}36) = 2{\cdot}5$ times as high at follow-up as at the initial observation; for the placebo, the effect is $\exp(0{\cdot}55) = 1{\cdot}7$. Table IV also summarizes results for the cumulative logit models. The results are substantively the same for either link. Parameter estimates are larger for the cumulative logit. This is expected since effects for that link refer to a broader response range.

7. MEAN RESPONSE MODELS

The models discussed next, proposed by Koch *et al.*,[1] are simpler to interpret than logit models but are structurally more controversial. They require assignment of scores $\{a_1, \ldots, a_r\}$ to response categories. The models describe the mean response, which for occasion g and covariate $\mathbf{x}$ is

$$M_g(\mathbf{x}) = \Sigma_k a_k \phi_{gk}(\mathbf{x}).$$

The models formulated in Table II also make sense for mean responses, substituting $M_g(\mathbf{x})$ for $L_{gk}(\mathbf{x})$ and deleting the k subscript from the α term. Unlike logit models, mean response models do not characterize marginal distributions in their entirety, but only through a measure of location.

Table V. Follow-up response independent of treatment, given initial response

Treatment	Initial	Follow-up Low	Medium	High	Total
Active	Low	30	15	0	45
	Medium	10	20	10	40
	High	0	5	10	15
	Total	40	40	20	
Placebo	Low	10	5	0	15
	Medium	10	20	10	40
	High	0	15	30	45
	Total	20	40	40	

Thus, for d.f. values in Table II to apply to mean response models, we must replace $r-1$ by 1. For $r=2$, mean response models are equivalent to linear probability models. At each occasion, such models assume a linear influence of $\mathbf{X}$ on the probability of either response. This assumption is structurally unsound, since one can obtain probabilities outside the [0, 1] range. It is less problematic as r increases, if we regard the scores merely as approximations for a discrete version of an underlying continuous scale.

We fitted mean response models to Table I using WLS, with response scores $\{10, 25, 45, 75\}$ for time to falling asleep. There are four response means and four parameters for the interaction model, so it is saturated; the fitted marginal means equal the observed ones. The initial means were 50·0 for the active drug group and 50·3 for the placebo, and the difference in means between the initial observation and the follow-up was 22·2 for the active drug and 13·0 for the placebo. The difference between these differences of means equals 9·2 (ase = 3·0), indicating a significantly greater change for the active drug.

8. COMPARING TREATMENT EFFECTS, CONTROLLING FOR INITIAL RESPONSE

Let us consider further the case, illustrated by Table I, of $d=2$ occasions. At each level of initial response, suppose the *conditional* distribution for follow-up response is identical for active and placebo treatments. Then if the marginal distributions for initial response are identical for both treatments, the follow-up marginal distributions are also identical. If the initial marginal distributions are not identical, however, the difference (on some scale) between the follow-up and initial marginal distributions may differ for active and placebo treatments. The artificial data in Table V illustrate this point. For the active treatment, the follow-up distribution is stochastically higher than the initial distribution; the reverse is true for the placebo treatment, even though the conditional distribution for follow-up response is identical for the two treatments.

Though models for marginal distributions can be useful for describing longitudinal effects when the initial marginal distributions differ, they may not tell the whole story. It is also informative to construct *transitional* models, which describe subject-wise patterns of change. Let L_{ijk} denote the cumulative logit when the cutpoint for follow-up response is at category k, for treatment i with baseline (initial) observation j. Let $\{x_j\}$ be fixed scores for the baseline levels. The model

$$L_{ijk}=\alpha_k+\beta_i+\beta x_j \tag{8}$$

uses $\{\beta_i\}$ to compare follow-up distributions, controlling for baseline observation. This is an analogue of an analysis of covariance model, in which the response and covariate are ordinal

rather than continuous. The transitional approach is less straightforward when responses occur at more than two occasions.

The joint distribution of initial and follow-up responses modelled in (8) is much more sparse than the marginal distributions. For instance, Table I has two '0' counts and four '1' counts, whereas the smallest marginal count is 11. Hence, we recommend using ML to fit model (8). Because of sparseness, goodness-of-fit statistics give only rough indices of quality of fit, though they are more adequate for comparing models. We report values of the likelihood-ratio goodness-of-fit statistic, denoted by G^2.

Applying model (8) to Table I with cumulative logit link and scores $\{10, 25, 45, 75\}$ for time to falling asleep, we obtain a likelihood-ratio statistic of $G^2 = 34{\cdot}8$ based on d.f. $= 19$. Adding a $\gamma_i x_j$ interaction term gives only a slightly better fit, with $G^2 = 30{\cdot}8$ based on d.f. $= 18$. Inspection of Table I reveals that for the first two baseline levels, the two treatments have similar sample distributions of time to fall asleep at the follow-up; at higher baseline levels, the active treatment seems more successful than the placebo.

We next fitted the interaction model

$$L_{ijk} = \alpha_k + \beta_i + \delta_j + \gamma_{ij}$$

in which $\beta_i = \delta_j = 0$ for all i and j, and $(\gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{14}, \gamma_{21}\, \gamma_{22}, \gamma_{23}, \gamma_{24}) = (\tau, \tau, \lambda + \sigma, \lambda, \tau, \tau, \sigma, 0)$. With this parameterization, there are four identical follow-up distributions, for active and placebo treatments at the first two levels of initial response. The difference between follow-up distributions for active and placebo treatments at the highest two levels of initial response is identical (λ). This model has ML estimates $\hat{\tau} = 3{\cdot}17$ (ase $= 0{\cdot}38$), $\hat{\lambda} = 1{\cdot}30$ (ase $= 0{\cdot}29$), and $\hat{\sigma} = 1{\cdot}16$ (ase $= 0{\cdot}29$). At the two highest levels for initial response, the odds that time to falling asleep is below any fixed level is estimated to be $\exp(1{\cdot}30) = 3{\cdot}7$ times higher for the active treatment than for the placebo. At those levels, the ratio $1{\cdot}30/0{\cdot}29 = 4{\cdot}5$ gives strong evidence that follow-up time to falling asleep is lower for the active treatment than the placebo. This model has $G^2 = 23{\cdot}5$, based on d.f. $= 18$, with lack of fit in a couple of cells. Further fine tuning improves the fit somewhat, but the improvement is minor relative to the resulting complexity of the interpretation. Other links give similar results. Francom, Chuang and Landis[5] gave an alternative transitional model for Table I, a log-linear model. Ware, Lipsitz and Speizer[3] described a way of generalizing the cumulative logit transitional model when the number of occasions is greater than 2.

9. USE OF STATISTICAL COMPUTER PACKAGES FOR MODEL FITTING

For single-response models, the ML approach can be implemented in SAS using the supplementary procedure LOGIST. For instance, LOGIST can be used to fit transitional models (such as model (8)), but not models for marginal distributions.

The WLS approach can be implemented for all models discussed in this article using procedure CATMOD is SAS. Table VI presents SAS code (Version 6 for the PC) for fitting a cumulative logit model to the margins of Table I. RESPONSE CLOGITS is a new option that specifies cumulative logits. Users not having version 6 of SAS with the CLOGITS option can use a RESPONSE statement to construct cumulative logits at each setting of **x**, in an analogous way to that described in the next paragraph for adjacent-category logits. The MODEL statement gives the design matrix for the linear predictor relating the margins of INITIAL (initial time to fall asleep) and FOLLOW (follow-up time to fall asleep) to cutpoint parameters (the first three columns), a treatment effect (column 4), an occasion effect (column 5), and interaction (column 6).

Table VII presents SAS (CATMOD) code for the adjacent-category logit model. The RESPONSE statement constructs adjacent-category logits for the margins, using the (**A** log **Bp**)

Table VI. Using SAS to fit cumulative logit model to margins of Table I

```
INPUT TREAT $ INITIAL $ FOLLOW $ COUNT @;
IF COUNT=0 THEN COUNT=0.000001; CARDS;
A 1 1 7   A 1 2 4   A 1 3 1   A 1 4 0
A 2 1 11  A 2 2 5   A 2 3 2   A 2 4 2
A 3 1 13  A 3 2 23  A 3 3 3   A 3 4 1
A 4 1 9   A 4 2 17  A 4 3 13  A 4 4 8
P 1 1 7   P 1 2 4   P 1 3 2   P 1 4 1
P 2 1 14  P 2 2 5   P 2 3 1   P 2 4 0
P 3 1 6   P 3 2 9   P 3 3 18  P 3 4 2
P 4 1 4   P 4 2 11  P 4 3 14  P 4 4 22
PROC CATMOD ORDER=DATA; WEIGHT COUNT;
POPULATION TREAT;
RESPONSE CLOGITS;
MODEL INITIAL*FOLLOW=(
                    1 0 0 1 1 1,
                    0 1 0 1 1 1,
                    0 0 1 1 1 1,
                    1 0 0 1 0 0,
                    0 1 0 1 0 0,
                    0 0 1 1 0 0,
                    1 0 0 0 1 0,
                    0 1 0 0 1 0,
                    0 0 1 0 1 0,
                    1 0 0 0 0 0,
                    0 1 0 0 0 0,
                    0 0 1 0 0 0)
(1 2 3='CUTPOINTS', 4='TREAT', 5='OCCAS', 6='TREAT*OCCAS');
```

response form. Here, **p** has 32 elements, and **B** consists of a block diagonal matrix with two blocks, each having form

$$\begin{matrix}
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1
\end{matrix}$$

When applied to **p**, one block forms the eight marginal proportions for the active treatment, and the other block forms the marginal proportions for the placebo treatment. Similarly, the **A** matrix is a block diagonal matrix with two components, each having form

$$\begin{matrix}
1&-1&0&0&0&0&0&0\\
0&1&-1&0&0&0&0&0\\
0&0&1&-1&0&0&0&0\\
0&0&0&0&1&-1&0&0\\
0&0&0&0&0&1&-1&0\\
0&0&0&0&0&0&1&-1
\end{matrix}$$

Table VII. Using SAS to fit interaction model to margins of Table I, using adjacent-category logits

```
PROC CATMOD ORDER=DATA;  WEIGHT COUNT;
 POPULATION TREAT;
   RESPONSE 1  -1   0   0  0   0   0   0,
            0   1  -1   0  0   0   0   0,
            0   0   1  -1  0   0   0   0,
            0   0   0   0  1  -1   0   0,
            0   0   0   0  0   1  -1   0,
            0   0   0   0  0   0   1  -1  LOG
  1  1  1  1  0  0  0  0  0  0  0  0  0  0  0  0,
  0  0  0  0  1  1  1  1  0  0  0  0  0  0  0  0,
  0  0  0  0  0  0  0  0  1  1  1  1  0  0  0  0,
  0  0  0  0  0  0  0  0  0  0  0  0  1  1  1  1,
  1  0  0  0  1  0  0  0  1  0  0  0  1  0  0  0,
  0  1  0  0  0  1  0  0  0  1  0  0  0  1  0  0,
  0  0  1  0  0  0  1  0  0  0  1  0  0  0  1  0,
  0  0  0  1  0  0  0  1  0  0  0  1  0  0  0  1;
   MODEL INITIAL*FOLLOW=( 1  0  0  1  1  1,
                          0  1  0  1  1  1,
                          0  0  1  1  1  1,
                          1  0  0  1  0  0,
                          0  1  0  1  0  0,
                          0  0  1  1  0  0,
                          1  0  0  0  1  0,
                          0  1  0  0  1  0,
                          0  0  1  0  1  0,
                          1  0  0  0  0  0,
                          0  1  0  0  0  0,
                          0  0  1  0  0  0)
(1  2  3='CUTPOINTS', 4='TREAT', 5='OCCAS', 6='INTERACTION')/PRED;
```

Table VIII. Using SAS to fit interaction model to margins of Table I, using mean responses

```
PROC CATMOD  ORDER=DATA;  WEIGHT COUNT;
POPULATION TREAT;
   RESPONSE 10  10  10  10  25  25  25  25  45  45  45  45  75  75  75  75,
            10  25  45  75  10  25  45  75  10  25  45  75  10  25  45  75;
MODEL INITIAL*FOLLOW=(1 1 1 1,  1 1 0 0,  1 0 1 0,  1 0 0 0)
   (2='TREAT', 3='OCCAS', 4='INTERACTION')/PRED;
```

for producing the differences of logs corresponding to the six marginal logits for each treatment. The POPULATION TREAT statement requests that these components be applied separately at the levels of the treatment variable. Thus ($\mathbf{A}\log\mathbf{B}\pi$) represents the 12 marginal adjacent-category logits, three in each margin for each treatment. The MODEL statement uses the same design matrix as in the cumulative logit model.

The default response in CATMOD is the set of logits in which each category is paired with the final category. Instead of using a RESPONSE statement to construct adjacent-category logits, one can use this default by specifying the design matrix that gives the equivalent model.

Table VIII presents SAS (CATMOD) code for the mean response model. The RESPONSE statement in that table calculates mean responses in the margins, using scores (10, 25, 45, 75). The MODEL statement fits the mean response model with interaction.

The semi-parametric approach is not currently available in statistical computer packages, but Stram, Wei and Ware[4] provide a FORTRAN program (called MORC).

10. COMPARISON OF LINKS, AND ALTERNATIVE APPROACHES

The advantages and disadvantages of the various model links described in Sections 5–7 are similar to those for the corresponding models for a single response. See Agresti[21] (Chapter 11) for a discussion of these.

Of the logit models, the cumulative logit has the advantage of a certain invariance to response category choice. If a cumulative logit model holds for an underlying continuous response, it also holds for any categorical measurement of the response, with the same values for the effect parameters. For sample data, if the model fits well for a fixed set of response categories, it also tends to fit well when we combine sets of adjacent responses, with similar ML estimates of effect parameters. This is not true for adjacent-category logit models. Its corresponding log-linear model assumes a scoring of response categories in which adjacent categories are equally distant; if this holds for one set of categories, it will generally not hold when we combine some response categories. When there is an arbitrary rather than a fixed choice of response categories, interpretation of parameters may also be more natural for cumulative logit models, referring to the entire response scale regardless of the cutpoints. When there is a fixed set of responses, the adjacent-category logit is sometimes more useful, since it permits contrasts with pairs of response categories.

Positive features of mean response models include (1) the interface with standard regression modelling that occurs as r increases, so that the response is more nearly continuous, and (2) the simplicity of interpretation of effects that refer to differences of means. Though such models do not contain cutpoint nuisance parameters, this omission also has disadvantages. When the number of response categories r exceeds 2, special cases of the model do not correspond exactly to conditions such as marginal homogeneity or statistical independence of response and covariates. Simply modelling a measure of location of the response does not allow comparisons of entire response distributions, such as whether they are stochastically ordered.

We conclude by describing a more complex way of modelling repeated ordered categorical data – one that still focuses on marginal distributions but uses multivariate information. Let Y_{gh} denote the response for subject h at occasion g. We permit a separate response distribution for each subject. Letting L_{gkh} be the linear predictor, we consider the model

$$L_{gkh}=\alpha_k+\lambda_h+v_g, \tag{9}$$

where one could add a $\boldsymbol{\beta}'\mathbf{x}$ term to allow for covariates. For n subjects, this model refers to a $d\times r\times n$ table of marginal distributions. There is a single observation at each occasion × subject combination.

Model (9) differs from those discussed in Section 3, which refer to population change rather than subject-wise change. For instance, model (2),

$$L_{gk}=\alpha_k+\mu_g,$$

refers to the occasion × response marginal table, collapsed over subjects. The $\{\mu_g\}$ here differ from the $\{v_g\}$ in (9). For cumulative logits, we interpret

$$v_1-v_2=\text{logit}[P(Y_{1h}\leqslant k)]-\text{logit}[P(Y_{2h}\leqslant k)],$$

whereas

$$\mu_1 - \mu_2 = \text{logit}[P(Y_{1h} \leqslant k)] - \text{logit}[P(Y_{2i} \leqslant k)],$$

where subject h is randomly selected at occasion 1 and subject i is randomly selected at occasion 2 (that is, h and i are *independent* observations). The parameters do agree when there is marginal homogeneity. If all $v_g = 0$, then all $\mu_g = 0$, because of the marginal occasion × subject independence.

The distinction between testing marginal homogeneity by testing all $\mu_g = 0$ in (2) and testing all $v_g = 0$ in (9) parallels the distinction between two types of tests of marginal homogeneity discussed by Darroch[19] for nominal variables. The first case is similar to Bhapkar's approach of directly using the marginal distributions, whereas the second case approaches the data in the form used by Cochran–Mantel–Haenszel statistics.

Unfortunately, likelihood techniques are inapplicable to fitting model (9), since the number of parameters has the same order as the number of subjects. For $d = 2$ and no covariates, McCullagh[22] gave a WLS solution (see also McCullagh[23]). An interesting problem for future research is to develop ways of fitting models with random subject effects for multiple occasions, simultaneously including effects of covariates. There has been some work incorporating subjects into models as random effects in other contexts.[24–26]

ACKNOWLEDGEMENTS

I am grateful to Drs Christy Chuang, Stuart Lipsitz, Laurel Smith, James Ware, and two referees for comments on this manuscript. I thank Drs Francom, Chuang, and Landis for sending me their paper, and Dr W. Stanish for advice on SAS.

REFERENCES

1. Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H., Jr and Lehnen, R. G. 'A general methodology for the analysis of experiments with repeated measurement of categorical data', *Biometrics*, **33**, 133–158 (1977).
2. Landis, J. R., Miller, M. E., Davis, C. S. and Koch, G. G. 'Some general methods for the analysis of categorical data in longitudinal studies', *Statistics in Medicine*, **7**, 109–137 (1988).
3. Ware, J. H., Lipsitz, S. and Speizer, F. E. 'Issues in the analysis of repeated categorical outcomes', *Statistics in Medicine*, **7**, 95–107 (1988).
4. Stram, D. O., Wei, L. J. and Ware, J. H. 'Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates', *Journal of the American Statistical Association*, **83**, 631–637 (1988).
5. Francom, S. F., Chuang, C. and Landis, J. R. 'A log-linear model for ordinal data to characterize differential change among treatments', *Statistics in Medicine*, **8**, 571–582 (1989).
6. Liang, K. and Zeger, S. L. 'Longitudinal data analysis using generalized linear models', *Biometrika*, **73**, 13–22 (1986).
7. Thompson, R. and Baker, R. J. 'Composite link functions in generalized linear models', *Applied Statistics*, **30**, 125–131 (1981).
8. Goodman, L. A. 'The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds', *Biometrics*, **39**, 149–160 (1983).
9. McCullagh, P. 'Regression models for ordinal data (with discussion)', *Journal of the Royal Statistical Society, Series B*, **42**, 109–142 (1980).
10. Anderson, J. A. and Philips, P. R. 'Regression, discrimination and measurement models for ordered categorical variables', *Applied Statistics*, **30**, 22–31 (1988).
11. Goodman, L. A. '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', *Annals of Statistics*, **13**, 10–69 (1985).
12. Haber, M. 'Maximum likelihood methods for linear and log-linear models in categorical data', *Computational Statistics & Data Analysis*, **3**, 1–10 (1985).

13. Haber, M. 'Log-linear models for correlated marginal totals of a contingency table', *Communications in Statistics – Theory and Methods*, **14**, 2845–2856 (1985).
14. Aitchison, J. and Silvey, S. D. 'Maximum-likelihood estimation of parameters subject to restraints', *Annals of Mathematical Statistics*, **29**, 813–828 (1958).
15. Stanish, W. M., Gillings, D. B. and Koch, G. G. 'An application of multivariate ratio methods for the analysis of a longitudinal clinical trial with missing data', *Biometrics*, **34**, 305–317 (1978).
16. Woolson, R. F. and Clarke, W. R. 'Analysis of categorical incomplete longitudinal data', *Journal of the Royal Statistical Society, Series A*, **147**, 87–99 (1984).
17. Lipsitz, S. 'Methods for analyzing repeated categorical outcomes', Unpublished PhD dissertation, Harvard University, 1988.
18. Landis, J. R., Lepkowski, J. M., Davis, C. S. and Miller, M. E. 'Cumulative logit models for weighted data from complex sample surveys', Proceedings of the Social Statistics Section of the Americal Statistical Association, 1987.
19. Darroch, J. N. 'The Mantel–Haenszel test and tests of marginal symmetry; fixed-effects and mixed models for a categorical response', *International Statistical Review*, **49**, 285–307 (1981).
20. SAS *SAS/STAT Guide for Personal Computers, Version 6*, SAS Institute Inc., Cary, NC, 1987.
21. Agresti, A. *Analysis of Ordinal Categorical Data*, Wiley, New York, 1984.
22. McCullagh, P. 'A logistic model for paired comparisons with ordered categorical data', *Biometrika*, **64**, 449–453 (1977).
23. McCullagh, P. 'On the elimination of nuisance parameters in the proportional odds model', *Journal of the Royal Statistical Society, Series B*, **46**, 250–256 (1984).
24. Laird, N. M. and Ware, J. H. 'Random effects models for longitudinal data', *Biometrics*, **38**, 963–974 (1982).
25. Stiratelli, R., Laird, N. and Ware, J. 'Random effects models for serial observations with binary responses', *Biometrics*, **40**, 961–971 (1984).
26. Ware, J. H. 'Linear models for the analysis of longitudinal studies', *American Statistician*, **39**, 95–101 (1985).