Appl. Statist. (1990)
39, *No.* 2, *pp.* 205–217

Partial Proportional Odds Models for Ordinal Response Variables

By BERCEDIS PETERSON† and FRANK E. HARRELL, JR

Duke University Medical Center, Durham, USA

[Received July 1988. Final revision June 1989]

SUMMARY

The ordinal logistic regression model that McCullagh calls the proportional odds model is extended to models that allow non-proportional odds for a subset of the explanatory variables. The maximum likelihood method is used for estimation of parameters of general and restricted partial proportional odds models as well as for the derivation of Wald, Rao score and likelihood ratio tests. These tests assess association without assuming proportional odds and test proportional odds against various alternatives. Simulation results compare the score test for proportional odds with tests suggested by Koch, Amara and Singer that are based on a series of binary logistic models.

Keywords: Ordinal logistic regression; Ordinal response variables; Proportional odds; Rao's efficient score statistic

1. Introduction

Attempts to extend the logistic regression model for binary response variables to allow for ordinal response variables have often involved modelling cumulative logits, i.e. when the outcomes of Y are ordinal and are assigned the values $0, 1, \ldots, k$, cumulative probabilities can be defined by

$$C_{ij} = \Pr(Y \geqslant j | \mathbf{X}_i), \qquad i = 1, \ldots, n, \quad j = 1, \ldots, k,$$

allowing a logistic model to be written as

$$\ln\left(\frac{C_{ij}}{1 - C_{ij}}\right) = \alpha_j + \mathbf{X}_i'\beta, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, k, \tag{1}$$

where $\alpha_1 > \alpha_2 > \ldots > \alpha_k$. Researchers who have suggested this model in some form include Snell (1964), Walker and Duncan (1967), Williams and Grizzle (1972) and McCullagh (1980). j indexes the k possible cumulative probabilities obtained from using k cut-offs to dichotomize Y. The regression coefficient β_l for the lth explanatory variable X_l is the log-odds ratio for the Y by X_l association, controlling for the remaining explanatory variables. Since β_l does not depend on j, the model assumes that the relationship between X_l and a dichotomized Y does not depend on j, the point at which the dichotomization is made. McCullagh (1980) calls this assumption of

†*Address for correspondence*: Box 31217, Duke University Medical Center, Durham, NC 27710, USA.

 0035-9254/90/39205 $2.00

TABLE 1
Distribution of disease by smoking status

	Degree of coronary artery disease				
	No disease				*Very severe disease*
	0	*1*	*2*	*3*	*4*
Non-smoker	334	99	117	159	30
Smoker	350	307	345	481	67

identical odds ratios across the k cut-offs the proportional odds assumption, and he calls model (1) the proportional odds model. Ashby *et al.* (1986) present a goodness-of-fit test for the proportional odds model and develop a statistical policy for amalgamating adjacent categories.

Since examples of non-proportional odds are not difficult to find, there is a question about how frequently the proportional odds assumption holds. For example, non-proportional odds are seen in Table 1 in a coronary artery disease data set collected at Duke University Medical Center. Here, the four cumulative log-odds ratios calculated from a five-level ordinal measure of coronary artery disease and a dichotomous smoking status variable are 1.04 ± 0.10, 0.65 ± 0.09, 0.46 ± 0.10 and 0.07 ± 0.22 respectively, for $j = 1, \ldots, 4$. The log-odds ratio is the largest when the disease variable is dichotomized as 'no disease' versus 'at least some disease', and the log-odds ratio is the smallest when the dichotomization is 'less than most severe disease' versus 'most severe disease'. The log-odds ratios for the intermediate dichotomizations are ordered between these two extremes, suggesting a linear trend in the log-odds ratios.

One way of assessing the validity of the proportional odds assumption involves interpreting the ordinal response variable as the manifestation of an underlying, unobservable, continuous random variable. Using this approach the α parameters of model (1) are considered the category boundaries that define the levels of Y. The proportional odds assumption then implies that all observations have a common variance (scale) on the underlying continuum, and tests of the X–Y association are seen as tests of location on this continuum. In this context, McCullagh (1980) proposes a model that allows the subpopulations, indexed by i, to vary in location as well as variance on this implied underlying variable, i.e. in his model,

$$\ln\left(\frac{C_{ij}}{1 - C_{ij}}\right) = \frac{\alpha_j + \mathbf{X}_i'\boldsymbol{\beta}}{\tau_i}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, k, \tag{2}$$

$\mathbf{X}_i'\boldsymbol{\beta}$ is the location for the ith subpopulation and τ_i is its scale or variance. Testing the equality of the τ_i is equivalent to testing the proportional odds assumption. If the interpretation of this model is based on the assumption of an underlying, continuous random variable, then use of the model implies that the underlying variable has the logistic distribution with variance $\tau_i^2 \pi^2/3$ (Malik, 1985). An example using this model is given in Section 5.

This interpretation of the proportional odds model is based on the assumption that all observations have been assigned a value on the ordinal variable using the same

category boundaries on the underlying continuum. In other words, all observations share a common set of α_js. Farewell (1982) has developed a test for this assumption, and a data set which violates this assumption is given in Section 4.

Using a different strategy, with no reference to an underlying continuous variable, Koch *et al.* (1985) develop a two-stage method of estimation called functional asymptotic regression methodology (FARM), described by Imrey *et al.* (1981), which uses Wald tests and weighted least squares. In the first stage of this procedure separate maximum likelihood analyses are used to estimate each of the cumulative logits, $\ln\{\Pr(Y \geqslant j)/\Pr(Y < j)\}$, $j = 1, \ldots, k$, as functions of the same set of p explanatory variables; thus each analysis is a logistic regression using a binary response variable. The regression coefficients from these k models are concatenated into one vector $\hat{\boldsymbol{\xi}}$, of length $k(p+1)$, and the variance of $\hat{\boldsymbol{\xi}}$, $\hat{\mathbf{V}}(\hat{\boldsymbol{\xi}})$, is calculated. Proportional odds can be tested by using the Wald statistic

$$Q_C = \hat{\boldsymbol{\xi}}' \mathbf{C}' \{\mathbf{C}\, \hat{\mathbf{V}}(\hat{\boldsymbol{\xi}})\, \mathbf{C}'\}^{-1} \mathbf{C}\, \hat{\boldsymbol{\xi}} \tag{3}$$

where $\mathbf{C}$ is a contrast matrix of rank c chosen to test proportional odds for any subset of the p explanatory variables. If the proportional odds assumption is found to hold for any explanatory variable, then in the second stage of the FARM analysis new regression coefficients that take proportionality into account are estimated using weighted least squares to fit models constrained by $\boldsymbol{\xi} = \mathbf{Z}\boldsymbol{\gamma}$. $\mathbf{Z}$ is a constant matrix of full rank u and size $k(p+1) \times u$, and $\boldsymbol{\gamma}$ is a $u \times 1$ vector of unknown second-stage parameters to be estimated. The goodness of fit of this model is tested with

$$(\hat{\boldsymbol{\xi}} - \mathbf{Z}\hat{\boldsymbol{\gamma}})'\, \hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\xi}})\, (\hat{\boldsymbol{\xi}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}) \sim \chi^2_{k(p+1)-u}, \tag{4}$$

where a non-significant value indicates an adequate fit. An example of this technique is given in Section 4.

The model developed using statistic (4) permits partial proportional odds in that some explanatory variables may meet the proportional odds assumption, while others may not. However, statistic (3) does not allow the assumption of proportional odds to be tested for one set of variables while constraining another set to have proportional odds (i.e. by estimating fewer parameters in the beginning), as is possible in the maximum likelihood method to be presented in this paper. Furthermore, a maximum likelihood procedure will have more power in some situations than a procedure that uses Wald tests (Hauck and Donner, 1977).

In particular, this paper formulates 'partial proportional odds models' in which a subset of the explanatory variables is not assumed to have proportional odds. A maximum likelihood fit to these models can be obtained, or the models can be used as a basis for score tests of proportional odds without estimating the additional parameters. Tests for trends in the log-odds ratios are also developed. With the exception of model (7) in Section 2.2, all maximum likelihood procedures presented in the remainder of this paper can now be implemented with the LOGIST procedure of SAS (SAS Institute, 1988). As pointed out in Peterson and Harrell (1988), the necessary SAS code is quite simple and involves the use of two new LOGIST options and one new LOGIST statement. The SAS code for the examples presented in Sections 4–6 is available from the authors on request.

2. Partial Proportional Odds Models

2.1. *Unconstrained Partial Proportional Odds Model*

We assume that n independent random observations are sampled and that the responses of these observations on an ordinal variable Y are classified in $k+1$ categories with $Y = 0, 1, \ldots, k$. Thus, each observation has an independent multinomial distribution. The model suggested for the cumulative probabilities is

$$C_{ij} = \Pr(Y \geqslant j | \mathbf{X}_i) = \frac{1}{1 + \exp(-\alpha_j - \mathbf{X}_i'\boldsymbol{\beta} - \mathbf{T}_i'\boldsymbol{\gamma}_j)}, \qquad j = 1, \ldots, k, \qquad (5)$$

where $\mathbf{X}_i$ is a $p \times 1$ vector containing the values of observation i on the full set of p explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients associated with the p variables in $\mathbf{X}_i$ (the elements of $\boldsymbol{\beta}$ are denoted by β_l, $l = 1, \ldots, p$), $\mathbf{T}_i$ is a $q \times 1$ vector, $q \leqslant p$, containing the values of observation i on that subset of the p explanatory variables for which the proportional odds assumption either is not assumed or is to be tested, and $\boldsymbol{\gamma}_j$ is a $q \times 1$ vector of regression coefficients associated with the q variables in $\mathbf{T}_i$, so that $\mathbf{T}_i'\boldsymbol{\gamma}_j$ is an increment associated only with the jth cumulative logit, $j = 1, \ldots, k$, and $\boldsymbol{\gamma}_1 = \mathbf{0}$. The elements of $\boldsymbol{\gamma}_j$ are denoted by γ_{jl}, $l = 1, \ldots, q$.

If $\boldsymbol{\gamma}_j = \mathbf{0}$ for all j, then this model reduces to the proportional odds model. Thus a simultaneous test of the proportional odds assumption for the q variables in $\mathbf{T}$ is a test of the null hypothesis that $\boldsymbol{\gamma}_j = \mathbf{0}$ for all $j = 2, \ldots, k$. Since $\boldsymbol{\gamma}_1 = \mathbf{0}$, the model uses only $\alpha_1 + \mathbf{X}_i'\boldsymbol{\beta}$ to estimate the odds ratio associated with the dichotomization of Y into $Y = 0$ versus $Y > 0$, whereas the estimation of the odds ratios associated with the remaining cumulative probabilities involve incrementing $\alpha_j + \mathbf{X}_i'\boldsymbol{\beta}$ by $\mathbf{T}_i'\boldsymbol{\gamma}_j$.

The log-likelihood for model (5) is

$$L = \sum_{i=1}^{n} \sum_{j=0}^{k} I_{ij} \log\{\Pr(Y = j | \mathbf{X}_i)\} = \sum_{i=1}^{n} \sum_{j=0}^{k} I_{ij} \log P_{ij},$$

where I_{ij} is an indicator variable for observation i such that $I_{ij} = 1$ if $Y_i = j$ and $I_{ij} = 0$ if $Y_i \neq j$, $j = 0, \ldots, k$, and P_{ij} is

$$\begin{aligned} P_{i0} &= 1 - C_{i1}, && \text{if } Y_i = 0, \\ P_{ij} &= C_{ij} - C_{i,j+1}, && \text{if } 0 < Y_i < k, \\ P_{ik} &= C_{ik}, && \text{if } Y_i = k. \end{aligned}$$

One consideration in fitting model (5) is constraining these cumulative probabilities to be between zero and unity. This is dealt with by invoking Jennrich and Sampson's (1968) step-halving technique in Hartley's (1961) modified Gauss–Newton algorithm to ensure that the log-likelihood continues to increase. Among the many partial proportional odds models analysed so far with this technique, only a few required step halving, and step halving solved all problems encountered.

2.2. *Constrained Partial Proportional Odds Models*

In the coronary artery disease–smoking example given earlier it was noted that there appeared to be a linear relationship between the five log-odds ratios and j, the point at which Y was dichotomized. Although model (5) would require four γ_{jl}

parameters to deal with this particular situation, a model constraining the γ_{jl} to be linear in j would require only one additional parameter in the model. Further, if such a simplification were appropriate for all predictor variables not having proportional odds, then model (5) could be rewritten as

$$C_{ij} = \Pr(Y \geq j | \mathbf{X}_i) = \frac{1}{1 + \exp(-\alpha_j - \mathbf{X}_i'\boldsymbol{\beta} - \mathbf{T}_i'\boldsymbol{\gamma}\Gamma_j)}, \qquad j = 1, \ldots, k. \quad (6)$$

Here the Γ_j are fixed prespecified scalars and $\Gamma_1 = 0$. The new parameter $\boldsymbol{\gamma}$ is a vector of length q whose elements, denoted by γ_l, are unsubscripted by j. Although $\boldsymbol{\gamma}$ is not dependent on j, it is multiplied by the fixed scalar constant Γ_j in the calculation of the jth cumulative logit.

In the coronary artery disease–smoking status example where $k = 4$ and a linear trend in the log-odds ratios is expected, the analyst would specify $\Gamma_j = j - 1$. Thus the log-odds ratio associated with the first cumulative logit ($j = 1$) is simply β_l, while the log-odds ratios associated with the second–fourth cumulative logits are $\beta_l + \gamma_l$, $\beta_l + 2\gamma_l$ and $\beta_l + 3\gamma_l$ respectively. From this example it can be seen that the Γ_j constants can be used to constrain the log-odds ratios to have a specified relationship among themselves. This relationship need not be linear. For example, if $\Gamma_k = 1$ and all the remaining $\Gamma_j = 0$, this would imply a constant log-odds ratio across the first $k - 1$ cumulative probabilities, with a divergence from proportional odds only for the kth cumulative probability. It makes sense to use the constrained model only if $k > 2$.

Model (6) is appropriate only if all q variables require identical constraints. A model that permits the constraints to vary is

$$C_{ij} = \Pr(Y \geq j | \mathbf{X}_i) = \frac{1}{1 + \exp\{-\alpha_j - \mathbf{X}_i'\boldsymbol{\beta} - \mathbf{T}_i'(\operatorname{diag}\boldsymbol{\Gamma}_j)\boldsymbol{\gamma}\}}, \qquad j = 1, \ldots, k, \quad (7)$$

where $\boldsymbol{\Gamma}_j$ is $q \times 1$ and the lth element of $\boldsymbol{\Gamma}_j$ is a prespecified constant to be used with the lth predictor variable. This model does not allow a variable to have unconstrained non-proportional odds, as in model (5). The most general model would allow some variables to have proportional odds (no γ_{jl} or γ_l parameters needed), some variables to have unconstrained non-proportional odds (γ_{jl} parameters needed) and the remaining variables to have individualized constrained non-proportional odds (γ_l parameters needed).

3. Tests Based on Partial Proportional Odds Models

3.1. *Likelihood Ratio and Score Tests of Proportional Odds*

The likelihood ratio statistic can be used to test the global proportional odds assumption by comparing the log-likelihood from the proportional odds model, model (1), with the log-likelihood from model (5), (6) or (7) in which q of the p explanatory variables have non-proportional odds. Although the likelihood ratio test has the most desirable statistical properties compared with its competitors, it is costly to implement since it requires two maximizations of likelihood functions. Also, there is always the problem of numerical difficulties (divergence) in obtaining maximum

likelihood estimates for the full set of parameters. Because of the computational complexity of this statistic, Rao's efficient score statistic (Rao, 1947, 1973) was used to develop a test of proportional odds. The implementation of the score statistic requires maximization of the log-likelihood only under the null hypothesis of proportionality. Furthermore, Lee *et al.* (1983) found that the score statistic compares favourably with the likelihood ratio statistic in a similar context involving Cox's (1972) proportional hazards model.

Partitioning the parameter space θ into (ψ, λ), the score statistic tests hypotheses about ψ while letting λ contain the nuisance parameters for which maximum likelihood estimates are obtained (for discussion, see Cox and Hinkley (1974)). In model (5) the global proportional odds assumption that $\gamma_{jl} = 0$, $j = 2, \ldots, k$, $l = 1, \ldots, q$, is tested with a $q(k-1)$ degrees of freedom (DF) score statistic by letting ψ contain the γ_{jl} parameters and λ contain the α_j and β_l parameters. The proportional odds assumption for the lth explanatory variable is tested with a $(k-1)$-DF score statistic by letting λ contain the α_j and β_l parameters as before, but $\psi = (\gamma_{2l}\ \gamma_{3l} \ldots \gamma_{kl})'$. In models (6) and (7) a global test of proportional odds for q variables against constrained non-proportional odds alternatives is a test of $\gamma_1 = \gamma_2 = \ldots = \gamma_q = 0$ and thus has q DF. Tests of proportional odds for each variable separately against these same alternatives have one DF.

In the description of the score statistic above the implication is that proportional odds can be tested for q of the p explanatory variables during the fitting of a proportional odds model. However, it is also possible to divide these q variables into two groups of size q_1 and q_2 $(q_1 + q_2 = q)$ and then to fit a partial proportional odds model so that q_1 variables have non-proportional odds while score tests are provided for q_2 variables. This would be recommended if, when using the score tests described earlier, q_1 of the q variables were found to have non-proportional odds. The analyst would then fit these q_1 variables for non-proportional odds and retest the remaining q_2 variables. The generalization of the previous discussion to handle this possibility is straightforward, i.e. the vector of parameters, λ, for which a maximum likelihood fit is obtained can now contain γ_{jl} or γ_j parameters (corresponding to unconstrained and constrained proportional odds respectively) in addition to α_j and β_l parameters. The γ_l or γ_{jl} parameters in the model are indexed by $l = 1, \ldots, q_1$. The γ_l or γ_{jl} parameters out of the model for which score tests are provided are indexed by $l = q_1 + 1, \ldots, q$.

3.2. *Wald Tests*

In any model fitted with maximum likelihood, one-DF Wald tests can be calculated by dividing the estimated regression coefficients by their respective standard errors obtainable from the information matrix. Multiple-DF Wald tests are calculated by using quadratic forms. Several Wald tests related to proportional odds are available from the fit of models (5)–(7). If a variable X_l $(l = 1, \ldots, q_1)$ is fitted for unconstrained non-proportional odds, then a Wald test of the association of this variable with the response variable has k DF, since the null hypothesis is H_0: $\beta_l = 0$; $\gamma_{jl} = 0$, $j = 2, \ldots, k$. Likewise, if a variable has constrained non-proportional odds, the comparable two-DF null hypothesis of no association is H_0: $\beta_l = 0$, $\gamma_l = 0$. Furthermore, $(k-1)$-DF Wald tests of proportional odds can be calculated for all variables fitted for unconstrained non-proportional odds (H_0: $\gamma_{2l} = \gamma_{3l} = \ldots = \gamma_{kl}$). Likewise, for variables fitted with a constraint, one-DF Wald tests of proportional odds can be calculated (H_0: $\gamma_l = 0$).

3.3. *Goodness-of-fit Tests of a Constraint*

Although the score and Wald tests of $\gamma_l = 0$ for variable X_l described earlier are tests of whether there are non-proportional odds in the form of a specified constraint across $\gamma_{2l}, \gamma_{3l}, \ldots, \gamma_{kl}$, they should not be interpreted as tests of the adequacy of the constraint, i.e. tests of whether a single γ_l parameter fits the data as well as $k-1$ γ_{jl} parameters. Such a test can be obtained, however, by using the likelihood ratio test to compare the log-likelihoods of unconstrained and constrained models. This gives an approximate chi-square with $k-1-1=k-2$ DF. An approximation to this test for each of the predictor variables indexed by $l = q_1 + 1, \ldots, q$ can be obtained by taking the difference between the $(k-1)$-DF score statistic for proportional odds and the one-DF score statistic for the prespecified constraint (Lee *et al.*, 1983; Tolley, 1978). Both of these statistics have drawbacks. The likelihood ratio test requires two maximizations and presents more potential convergence problems. The statistic discussed by Lee *et al.* is much simpler to compute, but it fluctuates in its performance compared with the more reliable likelihood ratio test. For these reasons, we propose a score test.

A score test of the goodness of fit of the constrained partial proportional odds model for variable X_1, say, can be obtained as follows. Let λ contain the α_j, β_l and γ_{jl} or γ_l parameters in a model for which a maximum likelihood fit is desired. γ_1 for variable X_1 is included among these parameters. Let ψ contain the $k-1$ γ_{j1} parameters for variable X_1. Since both γ_1 and the $k-1$ γ_{j1}s are in (ψ, λ), the parameter space is overspecified, i.e. the $k-1$ possible departures from proportional odds for variable X_1 are represented by k parameters. The score test of the adequacy of the constraint is a test of $\gamma_{j1} = 0$, $j = 2, \ldots, k$, and has $k-2$ DF, since one DF is taken up by the γ_1 associated with the constraint in the model.

4. Application to Severity of Nausea Data

This example applies models (5) and (6) to a data set given in Farewell (1982). Farewell examines the proportional odds assumption for the data in Table 2 by testing the assumption of common category boundaries for the two treatment groups. This approach was discussed briefly in Section 1. Farewell rejects the assumption of proportional odds and concludes that different boundaries or intervals on the underlying continuum were used to classify the subjects. Models (5) and (6) ignore this interpretation and directly assess the equivalence of the $k-1$ cumulative log-odds ratios. These are given in Table 3. Although the four-DF global score test of proportional odds available from model (5) is not significant ($\chi^2_4 = 6.26$), the observed log-odds ratios

TABLE 2
Data on severity of nausea for patients receiving chemotherapy with and without cisplatin

	Severity of nausea						*Total*
	None	*Mild*		*Moderate*		*Severe*	
	0	*1*	*2*	*3*	*4*	*5*	
No cisplatin	43	39	13	22	15	29	161
Cisplatin	7	7	3	12	15	14	58

TABLE 3
Predicted log-odds ratios for the data in Table 2

Observed ± standard error (predicted from model (5))	*Predicted from model (6)*	*Predicted from model (1)*
0.98 ± 0.44	1.064	0.911
1.18 ± 0.34	1.064	0.911
1.24 ± 0.33	1.064	0.911
0.98 ± 0.32	1.064	0.911
0.37 ± 0.37	0.435	0.911

suggest a divergence from proportional odds for the last dichotomization. This hypothesis is tested in model (6) with a constraint specified by $\Gamma_5 = 1$ and the remaining $\Gamma_j = 0$. The one-DF score test for this constraint is 4.82 and its three-DF score test for goodness of fit is 1.43; together these two statistics indicate a good fit. Furthermore, while the proportional odds model has a log-likelihood of -374.75 and a one-DF Wald test of the treatment effect of 11.27, model (6) has a log-likelihood of -372.19 and a two-DF Wald test of the treatment effect of 15.11. Both model (6) with $\Gamma_5 = 1$ and the model suggested by Farewell estimate the same number of parameters and offer almost exactly the same degree of improvement over the proportional odds model. The difference is in their interpretation.

The Farewell data set can be used to illustrate the mechanics of the Koch *et al.* (1985) FARM procedure. The first step is to fit the preliminary models:

$$\ln\left(\frac{C_{ij}}{1 - C_{ij}}\right) = \alpha_j + \beta_j x_i, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, 5,$$

to obtain $\hat{\boldsymbol{\xi}} = (\hat{\alpha}_1\ \hat{\beta}_1\ \hat{\alpha}_2\ \hat{\beta}_2 \ldots \hat{\alpha}_5\ \hat{\beta}_5)'$. Wald statistic (3) is then used to test $\beta_1 = \beta_2 = \ldots = \beta_5$ with

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

The resulting four-DF chi-square value of 5.86 is slightly smaller than the score statistic from the maximum likelihood procedure. A second-stage model requiring $\boldsymbol{\xi} = \mathbf{Z}\boldsymbol{\gamma}$ and constraining the log-odds ratios as before can be fitted with

$$\mathbf{Z}' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and $\boldsymbol{\gamma} = (\alpha_1^*\ \alpha_2^* \ldots \alpha_5^*\ \beta^*\ \gamma^*)'$. The goodness of fit of this new seven-parameter model relative to the original 10-parameter model is tested with statistic (4) and has a value of 1.41 with three DF.

TABLE 4
Quality of right eye vision in men and women

	Vision quality				*Total*
	Highest			*Lowest*	
	0	*1*	*2*	*3*	
Men	1053	782	893	514	3242
Women	1976	2256	2456	789	7477

5. Application to McCullagh's Vision Quality Data

McCullagh (1980) applied model (2) to the data in Table 4 and concluded that the difference between the two groups can be described solely by a difference in their scale parameters, i.e. $\log(\hat{\tau}_1/\hat{\tau}_2) = 0.272$ with standard error 0.025. He notes that men have relatively more values in the two extreme categories than do women. As an alternative way of interpreting these data, model (5) can be used to examine the validity of the proportional odds assumption. We find a two-DF score test of proportional odds of 137.34 and note that the $k-1$ cumulative log-odds ratios for low quality vision in women compared with men are 0.29, 0.00 and -0.47. Fitting a saturated model (5), whose parameters β_1, $\beta_1 + \gamma_{11}$ and $\beta_1 + \gamma_{21}$ equal these three observed log-odds ratios, a three-DF Wald test of the sex effect of 136.01 is obtained. Although it is valid to interpret this data set in terms of non-proportional odds, McCullagh's interpretation in terms of unequal variances on an underlying continuum is more elegant and more incisive.

6. Coronary Artery Disease Example

Non-proportional odds is evident in a coronary artery disease data set collected at Duke University Medical Center. Here it is found that in a sample of 2289 heart patients three of the five examined risk factors for a five-level index of heart disease have considerable non-proportional odds. In particular, sex, age and a binary smoking status variable (SMOKE) have linearly decreasing log-odds ratios, while binary variables indicating the presence of hypertension (HYPER) and the presence of high cholesterol (HICHOL) have proportional odds. Results are given in Table 5. The three-DF difference in the log-likelihoods of the proportional odds model and the non-proportional odds model shows the improvement offered by the non-proportional odds model. Comparison of the two models' Wald statistics and parameter estimates is also instructive. For example, in the proportional odds model, the estimated log-odds ratio for the sex effect is 1.696, whereas in model (6) four linearly decreasing log-odds ratios are estimated from sex's β_l and γ_l parameters as 1.970, 1.612, 1.254 and 0.896. Finally, the score tests show the goodness of fit of model (6) for all five predictor variables.

The example illustrates the flexibility of the models in this paper. Models can be fitted in which only a subset of the predictor variables has proportional odds, and constraints can be put on variables with non-proportional odds to reduce the number of DF needed to describe their association with the dependent variable. Furthermore, the

TABLE 5
Models for risk factors for coronary artery disease†

Model (1): proportional odds model (5 parameters): −2 log L = 6124.41

	1-DF Wald test of association		*3-DF score test of proportional odds*
	χ^2	$\hat{\beta}_i \pm$ *standard error*	χ^2
AGE	219.22	0.070 ± 0.005	3.18‡
SEX	309.71	1.696 ± 0.097	44.13
SMOKE	49.43	0.630 ± 0.090	27.22
HICHOL	46.64	0.545 ± 0.080	1.01
HYPER	20.62	0.366 ± 0.081	3.07

Model (6): constrained non-proportional odds model with linear constraint on AGE, SEX and SMOKE (8 parameters): −2 log L = 6069.47

		Wald tests for association with disease		
	DF	*Wald* χ^2	$\hat{\beta}_i \pm$ *standard error*	$\hat{\gamma}_i \pm$ *standard error*
AGE	2	224.21	0.080 ± 0.006	−0.008 ± 0.003
SEX	2	326.08	1.970 ± 0.111	−0.358 ± 0.061
SMOKE	2	66.08	0.884 ± 0.109	−0.247 ± 0.057
HICHOL	1	43.82	0.531 ± 0.080	
HYPER	1	19.09	0.354 ± 0.081	

	Score tests	
	DF	χ^2
Goodness of fit of linear constraint on AGE	2	3.30
Goodness of fit of linear constraint on SEX	2	2.01
Goodness of fit of linear constraint on SMOKE	2	2.07
Test of proportional odds for HICHOL	3	1.84
Test of proportional odds for HYPER	3	1.25

† Disease is defined as $Y = 0$ if there is at most insignificant disease in all arteries, $Y = 1$ if one artery is 75% or more occluded, $Y = 2$ if two arteries are 75% or more occluded, $Y = 3$ if three arteries are 75% or more occluded and $Y = 4$ if the left main artery is 75% or more occluded regardless of occlusion in other arteries.
‡ Although this test is non-significant, when it was recalculated after fitting a constraint to the SEX and SMOKE variables, the chi-square value changed to 9.56.

models allow proportional odds to be tested during the fit of a proportional odds model or a partial proportional odds model, thus offering an advantage over the Koch *et al.* FARM procedure which requires the estimation of all the parameters in a total unconstrained non-proportional odds environment.

7. Simulation Results

Since both the score test of proportional odds and the FARM Wald test of proportional odds are less costly alternatives to the likelihood ratio test, it is of interest to compare the powers of these two tests with the power of the likelihood ratio test. Although such a comparison is not available yet, simulations were performed to compare the score and Wald tests with each other. The scenarios for which random observations were simulated were defined by specifying $\mathbf{X}_i$, $i = 1, \ldots, n$, and the

α_j, β_l and γ_{jl} parameters, and then using the SAS function RANUNI (SAS Institute, 1985) to generate the required uniform random numbers. The size of the test for any given scenario was estimated by setting all $\gamma_{jl} = 0$, while powers for this scenario were estimated by using at least one non-zero γ_{jl} but leaving the α_j and β_l parameters the same. Only categorical explanatory variables were used, and thus each simulated data set could be represented by an $s \times (k + 1)$ cross-tabulation table, s being the number of independent subpopulations defined by the p explanatory variables. The selection of the α_j, β_l and γ_{jl} regression parameters determines the expected cell probabilities in the $s \times (k + 1)$ table. Thus, the difference between the null and non-null cases is not only a difference in γ_{jl} parameters but also a difference in cell probabilities.

The simulations reveal that both tests often give blatantly erroneous results when the cross-tabulation table for the response variable by an explanatory variable contains empty cells at an inner value of Y, i.e. $1 < Y < k$. Less blatant, but still suspicious, results are occasionally obtained if the table suffers from a general sparseness of cell sizes. The score test also suffers if the number of observations at one of the levels of Y is small relative to the total sample size. Since the occurrence of invalid statistics seems completely determined by small expected cell probabilities, valid observed test sizes do not imply that the corresponding powers will also be valid. Likewise, too large or too small test sizes do not imply that powers will be invalid.

Through use of 'variance inflation factors' (Marquardt and Snee, 1975), the source of the problem with the score statistic was revealed to involve to some extent near singularity in the information matrix, although many problems were unexplained by singularities and more research is necessary. The most plausible source of error for the FARM Wald tests would be in an ill-conditioned $\mathbf{C}\,\hat{\mathbf{V}}(\hat{\boldsymbol{\xi}})\,\mathbf{C}'$ matrix, or at least in $\hat{\mathbf{V}}(\hat{\boldsymbol{\xi}})$, but here again a good indicator of ill conditioning has not yet been developed. Consequently, since no trustworthy indicator of a test statistic's validity is available, size and power comparisons of the score and FARM statistics should be regarded with caution when some cell sizes are small.

Several size and power comparisons are presented in Table 6 for designs having no obvious problems with sparse cell sizes. The score test may be slightly anticonservative in that the simulated sizes are always greater than the nominal α levels. These differences, however, are not large. The difference in power for the two tests depends on both the nominal α level and the design. For example, in design 1 the difference in power at $\alpha = 0.01$ is much greater than at $\alpha = 0.10$, whereas in designs 2–5 the difference in power does not seem to depend on α. In all designs the score test never has less power than the FARM Wald test and sometimes has considerably more. No results are available to indicate under what conditions the score test has more power than the FARM test.

8. Discussion

This paper presents score tests of proportional odds as well as prediction models which allow non-proportional odds. The models offer a versatile method of analysing ordinal response variables, since the explanatory variables may have either proportional odds (one parameter for each covariate), unconstrained non-proportional odds (k parameters for each covariate) or constrained non-proportional odds (via a prespecified trend in log-odds ratios). The score tests are computationally inexpensive and allow proportional odds to be tested against a wide variety of non-proportional

TABLE 6
Size and power of several tests of proportional odds†

	Size Nominal α levels				*Power* Nominal α levels			
	0.01	*0.025*	*0.05*	*0.10*	*0.01*	*0.025*	*0.05*	*0.10*
Design 1‡								
FARM	0.004	0.016	0.048	0.110	0.520	0.702	0.794	0.898
Score	0.016	0.028	0.064	0.124	0.666	0.786	0.846	0.918
Design 2§								
FARM	0.026	0.042	0.056	0.102	0.174	0.300	0.424	0.596
Score	0.028	0.048	0.072	0.128	0.266	0.402	0.542	0.726
Design 3§§								
FARM	0.002	0.008	0.022	0.064	0.154	0.266	0.388	0.496
Score	0.020	0.034	0.070	0.126	0.272	0.396	0.490	0.622
*Design 4**								
FARM	0.012	0.032	0.060	0.118	0.180	0.266	0.372	0.494
Score	0.014	0.030	0.060	0.128	0.188	0.288	0.404	0.538
*Design 5***								
FARM	0.004	0.012	0.046	0.082	0.136	0.210	0.278	0.392
Score	0.022	0.038	0.058	0.124	0.230	0.330	0.430	0.546

† Each entry in the table is based on 500 samples.
‡ One-DF test of proportional odds: for sizes, the scenario evident in the 'final model (proportional odds)' of Table 3 of Koch *et al.* (1985) was simulated. For powers, $\gamma_{21} = 0$ was changed to $\gamma_{21} = 0.5$.
§ Eight-DF global test of proportional odds: for powers, we simulated the scenario obtained by fitting model (5) to the data in example 2 of Koch *et al.* (1985), with only smoking status having non-proportional odds. For sizes, all γ_{jl} parameters were set to zero.
§§ 16-DF global test of proportional odds: a 10-level dependent variable is regressed on two crossed binary predictor variables with $n = 100$ in each cell. For sizes, the α_j values were 3.18, 2.20, 1.39, 0.62, 0.00, −0.62, −1.39, −2.20 and −3.18 and both β_l values were 0.3. For powers, all 16 γ_{jl} were non-zero.
* 10-DF global test of proportional odds: a four-level dependent variable is regressed on five completely crossed binary predictor variables so that 32 subpopulations are defined with $n = 10$ each. For sizes, the α_j values were 0.405, −0.847 and −2.20 and all five β_l values were 0.5. For powers, $\gamma_{25} = 0.5$ and the remaining nine γ_{jl} were zero.
** Two-DF test of goodness of fit of a linear constraint: a five-level response variable is regressed on a binary predictor variable that defines two subpopulations of $n = 60$ each. For sizes, the α_j values were 0.69, −0.34, −1.61 and −2.4, $\beta_1 = 0.1$ and the γ_{j1} values were 0.1, 0.2 and 0.3. For powers, the γ_{j1} were changed to 0.4, 0.7 and 0.1.

odds alternatives. Since non-proportional odds are common, the models and tests presented here may be useful in the analysis of ordinal response variables.

Acknowledgements

This work was supported by research grants HS-03834 and HS-04873 from the National Center for Health Services Research, research grants HL-17670 and HL-36587 from the National Heart, Lung and Blood Institute, training grant LM-07003 and grant LM-03373 from the National Library of Medicine.

References

Ashby, D., Pocock, S. J. and Shaper, A. G. (1986) Ordered polytomous regression: an example relating serum biochemistry and haematology to alcohol consumption. *Appl. Statist.*, **35**, 289–301.

Cox, D. R. (1972) Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B, **34**, 187–220.
Cox, D. R. and Hinkley, D. V. (1974) *Theoretical Statistics*. London: Chapman and Hall.
Farewell, V. T. (1982) A note on regression analysis of ordinal data with variability of classification. *Biometrika*, **69**, 533–538.
Hartley, H. O. (1961) The modified Gauss–Newton method for fitting of non-linear regression functions by least squares. *Technometrics*, **3**, 269–280.
Hauck, Jr, W. W. and Donner, A. (1977) Wald's tests applied to hypotheses in logit analysis. *J. Am. Statist. Ass.*, **72**, 851–853.
Imrey, P. B., Koch, G. G. and Stokes, M. E. (1981) Categorical data analysis: some reflections on the log linear model and logistic regression. *Int. Statist. Rev.*, **49**, 265–283.
Jennrich, R. I. and Sampson, P. F. (1968) Application of stepwise regression to non-linear estimation. *Technometrics*, **10**, 63–72.
Koch, G. G., Amara, I. A. and Singer, J. M. (1985) A two-stage procedure for the analysis of ordinal categorical data. In *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences* (ed. P. K. Sen), pp. 357–387. Amsterdam: North-Holland.
Lee, K. L., Harrell, F. E., Tolley, H. D. and Rosati, R. A. (1983) A comparison of test statistics for assessing the effects of concomitant variables in survival analysis. *Biometrics*, **34**, 341–350.
Malik, H. J. (1985) Logistic distribution. In *Encyclopaedia of Statistical Sciences* (eds S. Kotz and N. L. Johnson), vol. 5, pp. 123–128. New York: Wiley.
Marquardt, D. W. and Snee, R. (1975) Ridge regression in practice. *Am. Statistn*, **29**, 3–20.
McCullagh, P. (1980) Regression models for ordinal data (with discussion). *J. R. Statist. Soc.* B, **42**, 109–142.
Peterson, B. L. and Harrell, Jr, F. E. (1988) Partial proportional odds models and the LOGIST procedure. *Proc. 13th A. Conf. SAS Users Group International, Orlando*, pp. 952–956. Cary: SAS Institute.
Rao, C. R. (1947) Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proc. Camb. Phil. Soc.*, **44**, 50–57.
——— (1973) *Linear Statistical Inference and its Applications*, 2nd edn. New York: Wiley.
SAS Institute (1985) *SAS User's Guide: Basics, Version 5 Edition*. Cary: SAS Institute.
——— (1988) Changes and enhancements to the SAS system, release 5.18, under OS and CMS. *SAS Technical Report P-175*. SAS Institute, Cary.
Snell, E. J. (1964) A scaling procedure for ordered categorical data. *Biometrics*, **20**, 592–607.
Tolley, H. D. (1978) A non-parametric test for composite hypotheses in survival analysis. *Ann. Inst. Statist. Math.* A, **30**, 281–295.
Walker, S. H. and Duncan, D. B. (1967) Estimation of the probability of an event as a function of several independent variables. *Biometrika*, **54**, 167–178.
Williams, D. and Grizzle, J. E. (1972) Analysis of contingency tables having ordered response categories. *J. Am. Statist. Ass.*, **67**, 55–63.