Scientific Research

Log-Link Regression Models for Ordinal Responses

Christopher L. Blizzard[1], Stephen J. Quinn[2], Jana D. Canary[1], David W. Hosmer[3]

[1]Menzies Research Institute Tasmania, University of Tasmania, Hobart, Australia
[2]Flinders Clinical Effectiveness, Flinders University, Adelaide, Australia
[3]Department of Public Health, University of Massachusetts, Amherst, USA
Email: Leigh.Blizzard@utas.edu.au, steve.quinn@flinders.edu.au, Jana.Canary@utas.edu.au, hosmer@schoolph.umass.edu

ABSTRACT

The adjacent-categories, continuation-ratio and proportional odds logit-link regression models provide useful extensions of the multinomial logistic model to ordinal response data. We propose fitting these models with a logarithmic link to allow estimation of different forms of the risk ratio. Each of the resulting ordinal response log-link models is a constrained version of the log multinomial model, the log-link counterpart of the multinomial logistic model. These models can be estimated using software that allows the user to specify the log likelihood as the objective function to be maximized and to impose constraints on the parameter estimates. In example data with a dichotomous covariate, the unconstrained models produced valid coefficient estimates and standard errors, and the constrained models produced plausible results. Models with a single continuous covariate performed well in data simulations, with low bias and mean squared error on average and appropriate confidence interval coverage in admissible solutions. In an application to real data, practical aspects of the fitting of the models are investigated. We conclude that it is feasible to obtain adjusted estimates of the risk ratio for ordinal outcome data.

1. Introduction

Several logit-link regression models have been proposed to deal with ordered categorical response data. Three of these are the adjacent categories model [1], the continuation-ratio model [2], and the cumulative odds model [3]. The last is referred to also as the proportional odds model [4]. The basis of each of these models is the discrete choice model [5] for nominal categorical outcomes that are also termed the multinomial logistic regression model [6].

The purpose of this paper is to investigate the practicality of fitting the ordinal models with a logarithmic link in place of the logit link. We refer to the resulting models as the adjacent categories (AC) probability model, the continuation-ratio (CR) probability model, and the proportional probability (PP) model. Each is a constrained form of the log multinomial model [7], the log-link counterpart of the multinomial logistic model. The ordinal log-link models make it possible to directly estimate different but related forms of the risk ratio in prospective studies and the prevalence ratio in cross-sectional studies, overcoming thereby a limitation of logit-link models.

Epidemiological research is grounded largely in assessment of average risk, and in that field the worth of the odds ratio as a measure of effect has long been questioned [8,9] particularly for prospective [10] and cross-sectional [11] data.

To describe the log-link models for ordinal data, we have adapted specialist terminology used for ordinal logistic models. Several authors [6,12,13] have distinguished "forwards" and "backwards" versions of the CR logit-link model, with the outcome categories taken in reverse order in the "backwards" version. For proportional odds models, O'Connell [14] distinguished between an "ascending" version for lower-ordered categories versus higher categories, and a "descending" version for higher-ordered categories versus lower categories. Accordingly, we distinguish "forwards-ascending" and "forwards-descending" versions of the AC probability model and the PP model. The two versions of each model produce coefficients that differ both in sign and magnitude. For the CR probability model, it is necessary to additionally distinguish "backwards-ascending" and "backwards-descending" versions because the four possible versions each produce a different set of estimates. For brevity, we

focus in what follows on the "forwards-descending" version of each model. The likelihoods of all versions are provided in Supplementary Materials that are available from the authors.

The paper is organized as follows. We describe and estimate with example data the AC probability model in Section 2, the CR probability model in Section 3, and the PP model in Section 4. Three issues in fitting these models are briefly surveyed in Section 5. The results of a simulation study of the performance of the three models are summarized in Section 6. An application to real data is given in Section 7, and the implications are summarized in Section 8.

2. The Adjacent Categories Probability Model

2.1. Log Multinomial Model

Consider an ordinal response variable $Y = 1, 2, \cdots, J$ with J ordered levels. Assume there are n independent observations of Y and of K non-constant covariates $X_1, X_2, \cdots, X_K$ and denote the observed data as $(y_i, \boldsymbol{x}_i)$ for $i = 1, 2, \cdots, n$ where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iK})'$. Denote the joint probabilities of occurrence of each of the levels of Y as:

$$\Pr(Y_i = j) = \mu_{ij}, \ i = 1, 2, \cdots, n; \ j = 1, 2, 3, \cdots, J$$

A requirement of a probability model is that $\sum_{j=1}^{J} \mu_{ij} = 1$, which identifies the probability of one category (say $j = \varphi$) because $\mu_{i\varphi} = 1 - \sum_{j \neq \varphi} \mu_{ij}$. For an ordinal outcome, the most compelling choices for the identified category are the first $(j = 1)$ or last $(j = J)$. In what follows, we consider a model in which the first outcome category is the identified category.

Assume that the probabilities μ_{ij} depend on the observed values of the covariates, and have the exponential form $\mu_j(\boldsymbol{x}_i) = \exp(\alpha_{j0} + \boldsymbol{x}_i'\boldsymbol{\alpha}_j)$ where α_{j0} and $\boldsymbol{\alpha}_j' = (\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jK})$ are parameters to be estimated. The log multinomial model for the *final $J-1$ outcomes* is:

$$\Pr(Y_i = j | \boldsymbol{x}_i) = \mu_j(\boldsymbol{x}_i) = \exp(\alpha_{j0} + \boldsymbol{x}_i'\boldsymbol{\alpha}_j) \quad (1)$$

for $i = 1, 2, \cdots, n$ and $j = 2, 3, \cdots, J$ where $\alpha_{1k} = 0$ for $k = 0, 1, 2, \cdots, K$ and hence $\exp(\alpha_{10} + \boldsymbol{x}_i'\boldsymbol{\alpha}_1) = 1$. The linear predictor is:

$$\alpha_{j0} + \boldsymbol{x}_i'\boldsymbol{\alpha}_j = \alpha_{j0} + \alpha_{j1}x_{i1} + \alpha_{j2}x_{i2} + \cdots + \alpha_{jK}x_{iK}$$

The likelihood and log likelihood of the data under this model are given in Supplementary Materials. The model can be fitted with software that provides a procedure for maximizing the log likelihood with respect to the $(J-1) \times (K+1)$ parameters α_{jk} for $j = 2, 3, \cdots, J$ and $k = 0, 1, 2, \cdots, K$.

Example data with $J = 3$ ordered outcomes and a single $(K = 1)$ dichotomous study factor are presented in **Table 1**. Armstrong and Sloan [15] used these data to demonstrate logit-link ordinal regression models.

For the example data, the log multinomial model for the final $J - 1 = 2$ outcomes involves estimation of the joint probability $\Pr(Y_i = 2 | \boldsymbol{x}_i) = \mu_2(\boldsymbol{x}_i)$ of the "Mild" outcome among all subjects, and the joint probability $\Pr(Y_i = 3 | \boldsymbol{x}_i) = \mu_3(\boldsymbol{x}_i)$ of the "Severe" outcome among all subjects.

The results of estimating the model are shown at left in panel A of **Table 3**. The baseline risk estimates are $\exp(\hat{\alpha}_{20}) = \exp(-1.897) = 0.15$ for the "Mild" outcome and $\exp(\hat{\alpha}_{30}) = \exp(-2.996) = 0.05$ for the "Severe" outcome, and the relative risk estimates are $\widehat{RR}_2 = \exp(\hat{\alpha}_{21}) = \exp(0.288) = 1.33$ for the "Mild" outcome and $\widehat{RR}_3 = \exp(\hat{\alpha}_{31}) = \exp(0.693) = 2.00$ for the "Severe" outcome. The estimates can be verified from the data in **Table 2**, and the estimated standard errors are

Table 1. Hypothetical example of ordinal response data.

Exposed	None	Mild	Severe	Total
Yes	70	20	10	100
No	80	15	5	100

Table 2. Component tables for forwards-descending ordinal probability models.

A. Joint probabilities					
Exp.	Not mild	Mild	Total	Not severe	Severe	Total
Yes	80	20	100	90	10	100
No	85	15	100	95	5	100
	$RR_2 = \dfrac{20/100}{15/100} = 1.33$			$RR_3 = \dfrac{10/100}{5/100} = 2.00$		

B. Conditional probabilities					
Exp.	None	Mild or Severe	Total	Mild	Severe	Total
Yes	70	30	100	20	10	30
No	80	20	100	15	5	20
	$RR_{Cond2} = \dfrac{30/100}{20/100} = 1.50$			$RR_{Cond3} = \dfrac{10/30}{5/20} = 1.33$		

C. Cumulative probabilities					
Exp.	None	Mild or Severe	Total	None or Mild	Severe	Total
Yes	70	30	100	90	10	100
No	80	20	100	95	5	100
	$RR_{Cum2} = \dfrac{30/100}{20/100} = 1.50$			$RR_{Cum3} = \dfrac{10/100}{5/100} = 2.00$		

identical to the values that can be calculated using a linear approximation to the variance of the logarithm of the risk or relative risk [16].

2.2. Forwards-Descending AC Probability Model

The particular assumption of the AC probability model is that the joint probabilities have a response to covariates that is log-linear in the coefficients and a multiple of category order. The *forwards-descending* AC probability model is:

$$\Pr\left(Y_i = j \mid x_i\right) = \mu_j^r\left(x_i\right) = \exp\left(\alpha_{j0}^r + x_i'\alpha_j^r\right) \quad (2)$$

for $i = 1, 2, \cdots, n$ and $j = 2, 3, \cdots, J$, and where the superscript r denotes a constrained estimate, and the intercepts α_{j0}^r and slopes:

$$\left\{\hat{\boldsymbol{\alpha}}_2^r, \hat{\boldsymbol{\alpha}}_3^r = 2 \times \hat{\boldsymbol{\alpha}}_2^r, \cdots, \hat{\boldsymbol{\alpha}}_J^r = \left(J-1\right) \times \hat{\boldsymbol{\alpha}}_2^r\right\}$$

comprise a set of $(J-1) \times (K+1)$ parameters to be estimated. This model can be estimated by fitting the log multinomial model (1) subject to $(J-2) \times K$ constraints on the slope parameters to require $\hat{\alpha}_j^r = (j-1) \times \hat{\alpha}_2^r$ for $j = 3, \cdots, J$.

For the example data, the ratio constraint is $\hat{\alpha}_{31}^r = (3-1) \times \hat{\alpha}_{21}^r$. The results of estimating the model are shown at right in panel A of **Table 3**. The constrained relative risk estimates are $\exp(0.321) = 1.38$ and $\exp(0.643) = 1.90$, which are plausible as fitted values to the unconstrained estimates ($\widehat{RR}_2 = 1.33$ and $\widehat{RR}_3 = 2.00$ respectively). The slope estimates in adjacent outcome categories increase by the additive factor $\hat{\alpha}_{21}^r = 0.321$ and, on the ratio scale, the relative risks increase by the multiplicative factor $\widehat{RR}_{AC} = \exp(0.321) = 1.38$.

For brevity, we refer to the estimate $\widehat{RR}_{AC} = 1.38$ as a "summary" relative risk when strictly it is not. It is instead the multiplicative factor relating relative risks in

Table 3. Results of fitting forwards-descending versions of three ordinal response log-link models.

Model and outcome	Unconstrained model			Constrained model		
	Coeff.	(SE)	P-value	Coeff.	(SE)	P-value
A. Joint probabilities	*Log multinomial model*			*AC probability model**		
Mild—all categories						
intercept	−1.897	(0.238)	<0.001	−1.918	(0.192)	<0.001
slope	0.288	(0.311)	0.355	0.321	(0.188)	0.088
Severe—all categories						
intercept	−2.996	(0.436)	<0.001	−2.963	(0.358)‡	<0.001
slope	0.693	(0.529)	0.190	0.643	(0.377)	0.088
Log likelihood	−141.469			−141.4781		0.892
B. Conditional probabilities				*CR probability model†*		
Mild/Severe—all categories						
intercept	−1.609	(0.200)	<0.001	−1.593	(0.184)	<0.0001
slope	0.405	(0.252)	0.107	0.379	(0.222)	0.088
Severe—Mild/Severe						
intercept	−1.386	(0.387)	<0.001	−1.451	(0.268)	<0.001
slope	0.288	(0.466)	0.536	0.379	(0.222)	0.088
Log likelihood	−141.469			−141.493		0.825
C. Cumulative probabilities				*PP model‡*		
Mild/Severe—all categories						
intercept	−1.609	(0.200)	<0.001	−1.609	(0.200)	<0.001
slope	0.405	(0.252)	0.107	0.405	(0.252)	0.107
Severe—all categories						
intercept	−2.996	(0.436)	<0.001	−2.813	(0.294)	<0.001
slope	0.693	(0.529)	0.190	0.405	(0.252)	0.107
Log likelihood	−141.469			−141.670		0.526

*Adjacent categories (AC) probability model estimated by fitting a log multinomial model for the final $J − 1 = 2$ outcomes subject to the constraint $\hat{\alpha}_{31}^c = (3-1) \times \hat{\alpha}_{21}^c$. †Continuation-ratio (CR) probability model estimated by fitting a forwards-descending conditional probability mode subject to the constraint $\hat{\beta}_{31}^c = \hat{\beta}_{21}^c$. ‡Proportional probability (PP) model estimated by fitting a forwards-descending cumulative probability model subject to the constraint $\hat{\gamma}_{31}^c = \hat{\gamma}_{21}^c$.

adjacent categories. It does not represent a common relative risk of advancing to the higher of adjacent categories, because that would require

$$\mu_j^r(\pmb{x}_i)/\mu_{j-1}^r(\pmb{x}_i) = \mu_{j-1}^r(\pmb{x}_i)/\mu_{j-2}^r(\pmb{x}_i)$$

for $j = 3, \cdots, J$ when, in general, these ratios are unequal and differ for each covariate holding.

2.3. Assessment of Loss of Model Fit

Three equivalent large-sample methods [17] are available to assess whether the constraints result in significant loss of model fit. Each has an asymptotic distribution that is chi-squared with degrees of freedom equal to the number of additional constraints imposed. Firstly, a test may be based on the likelihood ratio criterion [18]. Secondly, a test may be based on the efficient scores criterion [19]. Thirdly, a test may be based on the criterion proposed by Wald [20]. For a fully-constrained AC probability model in which the constraint is applied to each covariate, the null hypothesis for these tests is that $\pmb{\alpha}_j = (j-1) \times \pmb{\alpha}_2$ for $j = 2, 3, \cdots, J$. It is possible to apply the constraints to a subset of the covariates in a partially-constrained model [21], and to test a composite hypothesis. Readers are referred to Rao [17] for details.

For the example data, the AC probability model with log likelihood -141.478 is a constrained version of a log multinomial model with log likelihood -141.469. A likelihood ratio test with test statistic $G_o = -2\left[(-141.478) - (-141.469)\right] = 0.019$ and one degree of freedom gives $\Pr(\chi^2(1) > 0.019) = 0.891$. The score and Wald statistics are identical in the first four decimal places. The results suggest that the effect of exposure has been summarized into a single relative risk without significant loss of model fit.

3. The Continuation-Ratio Probability Model

3.1. Forwards-Descending Conditional Model

Consider ordinal outcome data arising from a hierarchical sequence of binary events in which the probability at each stage is conditional on having reached that stage. McCullagh and Nelder [22] provide an example of insemination of milch cows, and propose a model of the probability of impregnation of a cow at the jth attempt after $(j-1)$ unsuccessful attempts. This is the probability of terminating at the jth stage conditional on having reached that stage. O'Connell [14] provides an example of childhood proficiency in literacy in which mastery of any level of literacy requires mastery of all previous levels, and interest lies in the probability of success at a higher level. This is the probability of advancing beyond the $(j-1)^{st}$ stage conditional on hav-

ing reached that stage. Denote that probability as:

$$\Pr(Y_i \geq j | Y_i \geq j-1) = \nu_{ij}, \quad i = 1, 2, \cdots, n; \quad j = 2, 3, \cdots, J$$

Assume that the conditional probabilities ν_{ij} depend also upon the observed values of the covariates, and have the exponential form $\nu_j(\pmb{x}_i) = \exp(\beta_{j0} + \pmb{x}_i'\pmb{\beta}_j)$, where β_{j0} and $\pmb{\beta}_j = (\beta_{j1}, \beta_{j2}, \cdots, \beta_{jK})$ are parameters to be estimated. The *forwards-descending* conditional probability model is:

$$\Pr(Y_i \geq j | Y_i \geq j-1, \pmb{x}_i) = \nu_j(\pmb{x}_i) = \exp(\beta_{j0} + \pmb{x}_i'\pmb{\beta}_j) \quad (3)$$

for $i = 1, 2, \cdots, n$ and $j = 2, 3, \cdots, J$ where $\beta_{1k} = 0$ for $k = 0, 1, 2, \cdots, K$ and hence $\exp(\beta_{10} + \pmb{x}_i'\pmb{\beta}_1) = 1$. The linear predictor is:

$$\beta_{j0} + \pmb{x}_i'\pmb{\beta}_j = \beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \cdots + \beta_{jK}x_{iK}$$

The likelihood and log likelihood of the data under this model are given in Supplementary Materials. The model can be fitted with software that provides a procedure for maximising the log likelihood with respect to its $(J-1) \times (K+1)$ parameters β_{jk} for $j = 2, 3, \cdots, J$ and $k = 0, 1, 2, \cdots, K$. An alternative approach to fitting the model is also described in Supplementary Materials.

For the example data, the forwards-descending conditional probability model involves estimating the conditional probability $\Pr(Y_i \geq 2 | Y_i \geq 1, \pmb{x}_i) = \nu_2(\pmb{x}_i)$ of a "Mild" or "Severe" outcomes among all subjects, and the conditional probability $\Pr(Y_i = 3 | Y_i \geq 2, \pmb{x}_i) = \nu_3(\pmb{x}_i)$ of the "Severe" outcome among subjects with a "Mild" or "Severe" outcome. The component tables for the model are shown in panel B of **Table 2**.

The results of estimating the model are shown at left in panel B of **Table 3**. The relative conditional risk estimates are $\widehat{RR}_{Cond2} = \exp(0.405) = 1.50$ and $\widehat{RR}_{Cond3} = \exp(0.288) = 1.33$. Those estimates and their estimated standard errors can be verified from the data in **Table 2**.

3.2. Forwards-Descending CR Probability Model

The particular assumption of CR probability models is that the conditional probabilities have a response to covariates that is independent of the outcome category and log-linear in the coefficients.

The *forwards-descending* CR probability model is:

$$\Pr(Y_i \geq j | Y_i \geq j-1, \pmb{x}_i) = \nu_j^c(\pmb{x}_i) = \exp(\beta_{j0}^c + \pmb{x}_i'\pmb{\beta}_j^c) \quad (4)$$

for $i = 1, 2, \cdots, n$ and $j = 2, 3, \cdots, J$ where the superscript c denotes a constrained estimate, and intercepts β_{j0}^c and common slope coefficients $\pmb{\beta}_2^c = \pmb{\beta}_3^c = \cdots = \pmb{\beta}_J^c$ comprise a set of $(J-1) \times (K+1)$ parameters to be estimated of which only $(J-1) \times (2)$ are unique. This model can be estimated by fitting the conditional probability model (3) subject to $(J-2) \times K$ constraints on

the slope parameters to require $\hat{\boldsymbol{\beta}}_j^c = \hat{\boldsymbol{\beta}}_{j-1}^c$ for $j = 3, \cdots, J$. The superscript c is used here to distinguish this equality form of constraint from the ratio constraint used in the AC model. An alternative approach to fitting the model is described in Supplementary Materials.

For the example data, the required parameter constraint is $\hat{\beta}_{31}^c = \hat{\beta}_{21}^c$. The results of estimating the model are shown at right in panel B of **Table 3**. The common slope estimate is $\hat{\beta}_{21}^c = \hat{\beta}_{31}^c = 0.379$, and the common relative conditional risk estimate is $\widehat{RR}_{CR} = \exp(0.379) = 1.46$. It plausibly falls between the unconstrained estimates.

The CR probability model with log likelihood -141.493 is a constrained version of the conditional probability model with log likelihood -141.469. A likelihood ratio test with statistic $G_o = 0.049$ gives $\Pr(\chi^2(1) > 0.049) = 0.825$. A score test provides an identical result. A Wald test with statistic $W_o = 0.050$ gives $\Pr(\chi^2(1) > 0.050) = 0.824$.

4. The Proportional Probability Model

4.1. Forwards-Descending Cumulative Probability Model

One further way of partitioning the ordered response categories is to form $(J-1)$ cumulative "splits" of the outcome data between category 1 and all higher categories, category 1 or 2 and all higher categories, and so on. This prompts a model of the cumulative probability of falling in category j or a higher category:

$$\Pr(Y_i \geq j) = \sum_{\varphi=j}^{J} \left[\Pr(Y_i = \varphi) \right] = \pi_{ij}$$

for $i = 1, 2, \cdots, n$ and $j = 2, 3, \cdots, J$. Assume that the cumulative probabilities π_{ij} depend upon the observed values of the covariates, and have the exponential form

$$\pi_j(\boldsymbol{x}_i) = \exp(\gamma_{j0} + \boldsymbol{x}_i'\boldsymbol{\gamma}_j)$$

where γ_{j0} and $\boldsymbol{\gamma}_j' = (\gamma_{j1}, \gamma_{j2}, \cdots, \gamma_{jK})$ are parameters to be estimated.

The *forwards-descending* cumulative probability model is:

$$\Pr(Y_i \geq j | \boldsymbol{x}_i) = \pi_j(\boldsymbol{x}_i) = \exp(\gamma_{j0} + \boldsymbol{x}_i'\boldsymbol{\gamma}_j) \quad (5)$$

for $i = 1, 2, \cdots, n$ and $j = 2, 3, \cdots, J$ and where $\gamma_{1k} = 0$ for $k = 0, 1, 2, \cdots, K$, and hence $\exp(\gamma_{10} + \boldsymbol{x}_i'\boldsymbol{\gamma}_1) = 1$. The linear predictor is:

$$\gamma_{j0} + \boldsymbol{x}_i'\boldsymbol{\gamma}_j = \gamma_{j0} + \gamma_{j1}x_{i1} + \gamma_{j2}x_{i2} + \cdots + \gamma_{jK}x_{iK}$$

The likelihood and log likelihood of the data under this model are given in Supplementary Materials. The model can be fitted with software that provides a procedure for maximising the log likelihood with respect to the $(J-1)\times(K+1)$ parameters γ_{jk} for $j = 2, 3, \cdots, J$ and $k = 0, 1, 2, \cdots, K$.

For the example data, the forwards-descending cumu-

lative probability model involves estimating the cumulative probability $\Pr(Y_i \geq 2 | \boldsymbol{x}_i) = \pi_2(\boldsymbol{x}_i)$ of "Mild" or "Severe" outcomes, and the probability $\Pr(Y_i = 3 | \boldsymbol{x}_i) = \pi_3(\boldsymbol{x}_i)$ of the "Severe" outcome. The component tables for the model are shown in panel C of **Table 2**. The results of estimating the model are shown at left in panel C of **Table 3**. The relative cumulative risk estimates are $\widehat{RR}_{Cum2} = \exp(0.405) = 1.50$ and $\widehat{RR}_{Cum3} = \exp(0.693) = 2.00$. Those estimates and their estimated standard errors can be verified from the data in **Table 2**.

4.2. Forwards-Descending PP Model

The particular assumption of the PP model is that the cumulative probabilities have a response to covariates that is independent of the outcome category and log-linear in the coefficients. The *forwards-descending* PP model is:

$$\Pr(Y_i \geq j | \boldsymbol{x}_i) = \pi_j^c(\boldsymbol{x}_i) = \exp(\gamma_{j0}^c + \boldsymbol{x}_i'\boldsymbol{\gamma}_j^c) \quad (6)$$

where the intercepts γ_{j0}^c and common slope coefficient vectors $\boldsymbol{\gamma}_2^c = \boldsymbol{\gamma}_3^c = \cdots = \boldsymbol{\gamma}_J^c$ comprise a set of $(J-1)\times(K+1)$ parameters to be estimated of which only $(J-1)\times(2)$ are unique. This model can be estimated by fitting the cumulative probability model (5) subject to $(J-2)\times K$ constraints on the slope parameters to require $\hat{\boldsymbol{\gamma}}_j^c = \hat{\boldsymbol{\gamma}}_{j-1}^c$ for $j = 3, \cdots, J$.

For the example data, the required parameter constraint is $\hat{\gamma}_{31}^c = \hat{\gamma}_{21}^c$. The results of estimating the model are shown at right in panel C of **Table 3**. The common slope estimate is $\hat{\gamma}_{21}^c = \hat{\gamma}_{31}^c = 0.405$, and the common relative cumulative risk estimate is $\widehat{RR}_{PP} = \exp(0.405) = 1.50$.

The PP model with log likelihood -141.670 is a constrained version of the cumulative probability model with log likelihood -141.469. A likelihood ratio test with $G_o = 0.402$ gives $\Pr(\chi^2(1) > 0.402) = 0.526$. A score test with statistic $S_o = 0.397$ yields $\Pr(\chi^2(1) > 0.397) = 0.529$. A Wald test with statistic $W_o = 0.382$ gives $\Pr(\chi^2(1) > 0.382) = 0.537$.

4.3. Three Properties of a Fully-Constrained Cumulative Probability Model

Firstly, the constrained value of the largest fitted cumulative probability at each set of covariate values, $\pi_2^c(\boldsymbol{x}_i)$, is identical to its unconstrained value $\pi_2(\boldsymbol{x}_i)$. Comparing solutions to the constrained and unconstrained models in panel C of **Table 3**, we see that $\hat{\gamma}_{20}^c = \hat{\gamma}_{20} = -1.609$ and $\hat{\gamma}_{21}^c = \hat{\gamma}_{21} = 0.405$, and hence that $\pi_2^c(x_1) = \pi_2(x_1)$. This remarkable result is a property of a fully-constrained cumulative probability model.

Secondly, the PP model (6) can be fitted as a log multinomial model (1) estimated subject to an equality constraint on the coefficients of each covariate (see Supple-

mentary Materials). Hence the common slope estimate $\hat{\gamma}_{21}^c = \hat{\gamma}_{31}^c = 0.405$ from the fully constrained cumulative probability model can be seen as a summary estimate for the unconstrained slope estimates $\hat{\alpha}_{21} = 0.288$ from the log multinomial model. The PP model forces these joint probabilities to be equal.

Thirdly, in a fully-constrained model, the fitted probabilities $\hat{\pi}_j^c(x_i)$ are proportional in the sense that

$$\hat{\pi}_j^c\left(x_i^{(1)}\right)\Big/\hat{\pi}_j^c\left(x_i^{(2)}\right) = \hat{\pi}_{j-1}^c\left(x_i^{(1)}\right)\Big/\hat{\pi}_{j-1}^c\left(x_i^{(2)}\right)$$

and

$$\hat{\pi}_j^c\left(x_i^{(1)}\right)\Big/\hat{\pi}_{j-1}^c\left(x_i^{(1)}\right) = \hat{\pi}_j^c\left(x_i^{(2)}\right)\Big/\hat{\pi}_{j-1}^c\left(x_i^{(2)}\right)$$

for $j = 3, \cdots, J$ where $x_i^{(1)}$ and $x_i^{(2)}$ denote distinct covariate holdings. These ratio equalities justify the description of the fully-constrained cumulative probability model as a proportional probability (PP) model.

5. Issues in Fitting the Models

5.1. Starting Values

A careful choice of starting values is required for the cumulative probability models. This is a consequence of the need to evaluate terms of the form $\ln\left[\pi_j(x_i) - \pi_{j+1}(x_i)\right]$ in the log likelihood. As starting values we use the estimates from the common-slopes log multinomial model, with intercept values derived from the recursive relation given previously.

5.2. Non-Admissable Solutions

Solutions to a probability model are not admissible if any of the fitted probabilities exceed unity and if the sum of the fitted probabilities exceeds unity. The functional form of the three ordinal models does not constrain the fitted probabilities to take admissible values, and hence non-admissible solutions can occur in the extremes of covariate space if the model-based probabilities are very high.

5.3. Non-Convergence

The log likelihoods for each model contain terms that cannot be evaluated if the fitted values of the probabilities, or their sums in the AC model, reach or exceed unity. This may prevent convergence to a solution. For model-based simulated data, the simple expedient of requiring evaluation of a term in the log likelihood only if its coefficient is non-zero eliminates almost all problems with convergence. Converged but non-admissible solutions are produced instead. We take advantage of this to study non-admissible solutions in the next section.

6. Simulation Study

To confirm that the three ordinal log-link regression

models produce estimates that match their theoretical values on average and without excessive variation, and with estimated standard errors that provide appropriate confidence interval coverage, we repeatedly generated data from each probability model and fitted the ordinal models to each dataset. The simulations involved a single $(K = 1)$ uniformly-distributed covariate X and a response variable Y having $J = 3$ ordered levels with $\Pr(Y = 2 | x = 0) = 0.12$, $\Pr(Y = 3 | x = 0) = 0.04$ and $\Pr(Y = j | x = 1) = (1.1) \times \Pr(Y = j | x = 0)$. We drew samples of size $n = 200$ and $n = 500$, and with an adequate number of replications to produce at least 10,000 admissible solutions and at least 1000 non-admissible solutions in two settings. The settings used and methods of generating the data are described in the Supplementary Materials. The results are presented there as average percent relative bias, average percent mean squared error and average 95 percent confidence interval coverage. Additionally, the results of likelihood ratio, score and Wald tests of the constraint imposed in each model were recorded.

Each of the three ordinal log-link regression models produced estimates with minimal bias and mean squared error on average, and with 95 percent confidence interval coverage that was close to the nominal 95 percent. The summary model fit statistics were similar for each model, and the minor differences did not impart a clear advantage for any single model. Type 1 error rates were close to 5 percent but, particularly for the CR probability model, with a tendency to reject slightly too often. At larger values of the uniform covariate, for which the theoretical values from three probability models are most different, power to reject an incorrect model was moderate for each of the AC probability model and the PP model, and higher for the CR probability model. Bias and mean squared error were smaller, error rates on average were closer to 5 percent, and power in each setting was greater, in the larger datasets $(n = 500)$ than in the smaller datasets $(n = 200)$.

The preceding comments refer to solutions for which the fitted values are admissible for a probability model. In non-admissible solutions that were deliberately engineered by allowing the uniform covariate to take large values, model fit was less satisfactory because bias was relatively high and confidence interval coverage was greatly reduced, and the tendency for each model to reject too often a correctly fitted model was more pronounced.

7. Application to Real Data

The dataset consists of information on 189 births to women seen in the obstetrics clinic at Baystate Medical Center in Springfield, Massachusetts. Hosmer and Lemeshow [6] used these data to demonstrate ordinal logit-link regression models. The ordered outcome consists of

$J = 4$ categorizations of birth weight (≤2500 g, 2501 - 3000 g, 3001 - 3500 g and >3500 g). The principal study factor for this analysis is maternal smoking status during pregnancy. The data on birth weight category and maternal smoking status are shown in **Table 4**.

In respect of the ordering of outcome categories, we chose to retain the natural ordering of the birth weight categories from lightest $(j = 1)$ to heaviest $(j = 4)$ rather than reverse them as Hosmer and Lemeshow [6] did, and to fit forwards-descending models with the first category as the identified category. For log-link models, these decisions are important because reversing the ordering of the outcome categories and/or modeling them in ascending order results in different models of the data.

In respect of choice of modeling approach, the CR probability model assumes the birth weight categories are the outcomes of a sequence of binary events when they arose from arbitrary categorization of an underlying continuous response variable. The AC and PP models provide poorer descriptions of the data in **Table 4**, however. The relative risk estimates are $\widehat{RR}_2 = 1.13$, $\widehat{RR}_3 = 0.91$

and $\widehat{RR}_4 = 0.49$ for the 2501 - 3000 g, 3001 - 3500 g and >3500 g categories respectively. They suggest that the impact of maternal smoking is restricted mainly to the highest birth weight category, but the AC probability model requires the response to covariates to increase monotonically and multiplicatively with birth weight category whilst the PP model constrains it to be identical in each outcome category.

The results of fitting the three models to the birth weight data are summarized in **Table 5**. Estimation results are provided for a binary (0 = no, 1 = yes) covariate for maternal smoking (*smoker*). The common relative risk increase is $\widehat{RR}_{AC} = 0.88$, the summary relative conditional risk is $\widehat{RR}_{CR} = 0.81$, and the summary relative cumulative risk is $\widehat{RR}_{PP} = 0.80$. Also shown in **Table 5** are the results of adjusting for confounding factors including maternal age (*age*). In these models, only the coefficients of *smoker* are constrained. The adjustment has increased the estimated summary effect of maternal smoking in each model. The adjusted values are $\widehat{RR}_{AC} = 0.83$, $\widehat{RR}_{CR} = 0.76$ and $\widehat{RR}_{PP} = 0.74$. Reflecting

Table 4. Birth weight and maternal smoking during pregnancy.

Maternal smoking status	Birth weight (g)				Total
	≤2500	2501 - 3000	3001 - 3500	>3500	
Yes	30	16	17	11	74
No	29	22	29	35	115

Table 5. Estimation results for the binary covariate for maternal smoking status from three ordinal response log-link models applied to the low birth weight data.

	Unconstrained model			Constrained model		
	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)
AC model[†]						
Unadjusted	0.122 (0.293)	−0.093 (0.267)	−0.717 (0.312)[*]	−0.128 (0.055)[*]	−0.256 (0.109)[*]	−0.384 (0.164)[*]
Log likelihood		−255.486			−256.429 (P = 0.390)	
Adjusted[‡]	0.024 (0.3207)	−0.110 (0.312)	−1.04 (0.299)[*]	−0.191 (0.050)[*]	−0.381 (0.100)[*]	−0.572 (0.150)[*]
Log likelihood		−224.716			−226.806 (P = 0.124)	
CR model[§]						
Unadjusted	−0.229 (0.110)	−0.157 (0.130)	−0.331 (0.261)	−0.214 (0.081)[*]	−0.214 (0.081)[*]	−0.214 (0.081)[*]
Log likelihood		−255.486			−255.694 (P = 0.812)	
Adjusted[‡]	−0.301 (0.093)[*]	−0.191 (0.178)	−0.291 (0.260)	−0.279 (0.077)[*]	−0.279 (0.077)[*]	−0.279 (0.077)[*]
Log likelihood		−223.721			−223.872 (P = 0.860)	
PP model[¶]						
Unadjusted	−0.229 (0.110)[*]	−0.386 (0.171)[*]	−0.717 (0.312)[*]	−0.229 (0.110)[*]	−0.229 (0.110)[*]	−0.229 (0.110)[*]
Log likelihood		−255.486			−257.219 (P = 0.177)	
Adjusted[‡]	−0.301 (0.093)[*]	−0.491 (0.200)[*]	−0.782 (0.329)[*]	−0.301 (0.093)[*]	−0.301 (0.093)	−0.301 (0.093)
Log likelihood		−223.721			−225.036 (P = 0.269)	

[*]Denotes P < 0.05; [†]Adjacent categories probability model; [‡]Added covariates are linear predictors for maternal age (*age*) and weight of the mother at the last menstrual period (*lwt*), two binary indicators for *race* (1 = white, 2 = black, 3 = other), and binary indicators for presence of uterine irritability (*ui*) and history of premature labour (*ptd*); [§]Continuation-ratio probability model; [¶]Proportional probability model.

the pronounced impact of maternal smoking in the highest birth weight category, the AC probability model estimate is less extreme that those of the CR model, and the CR probability model estimate is less extreme than that of the PP model. To simplify matters for demonstration purposes, non-linearity in the scale of the continuous covariates (particularly *age*) and statistical interaction between covariates (including between *smoker* and *age*) have been ignored. The resultant models had 6 or more covariates including continuous variables, and produced converged and admissible solutions. Non-convergence was an occasional problem only, and occurred with very complex models that involved higher-powered polynomials of continuous variables and/or product terms for statistical interaction.

The likelihood ratio tests reveal that the constraints imposed to produce the summary estimates have resulted in substantially greater loss of model fit for the AC and PP models than for the CR model. If the subset $(n = 189)$ of data analyzed here is representative of all data collected in the birth weight study, the loss of model fit would be statistically significant if the full dataset was two (AC probability model), three (PP model) or 20 (CR probability model) times larger in size.

8. Discussion

In this paper, we proposed log-link alternatives to three logistic regression models for ordinal data, and demonstrated that it is possible to estimate these models with software that allows the user to specify the log likelihood as the objective function to be maximized and to impose constraints on the parameter estimates. In example data with a dichotomous covariate, the unconstrained models produced relative risk estimates and estimated standard errors that could be verified from the data, and the constrained estimates were credible approximations to the unconstrained estimates. Models with a single continuous covariate performed well in data simulations, with low bias and mean squared error on average and appropriate confidence interval coverage in admissible solutions. Finally, the models were successfully fitted to a complex real-world dataset with an ordinal outcome for which one of the models provided a questionable theoretical explanation but a superior practical description.

When the model-based probabilities were high, issues in fitting the ordinal response log-link regression models arose because the fitted probabilities are not restricted to values less than unity. The estimation algorithm may fail to converge or, if it does converge, the fitted values may not be admissible for a probability model. It is possible to write the algorithm in a manner that makes non-convergence a rare occurrence in model-based simulated data (less than 1 per 10,000 replications in our settings). The

advantage of doing so in simulated data is that the estimation algorithm converges to a non-admissible solution, and this makes it feasible to study them. We found that the estimated coefficients of the inadmissible solutions had much greater bias on average, with estimated standard errors that were too small, and 95% confidence interval coverage that was poorer in consequence. These findings caution against relying on the results of non-admissible solutions.

Each of the three ordinal response log-link models we considered is a constrained version of the log multinomial model. The constraints impose a simple linear relation on the slope coefficients, and it is possible to test whether the unconstrained model provides a better description of the data. We found that likelihood ratio, score and Wald tests of the constraints each had close to the correct type I error rates though they tended to be slightly conservative. The power of tests for each model to reject fits to data from any of the other ordinal response log-link models increased with the size of the maximum model-based probability, and was moderate to high when the maximum model-based probability was high.

In our estimation procedure, the user has flexibility to place linear constraints on the coefficients of all, or some, of the covariates. This makes it possible to fit "partially-constrained" models in which the linear constraint is applied to a subset only of the explanatory variables [21]. In the analyses of the birth weight data, which focused on the relationship between maternal smoking and birth weight outcome, it served no purpose to place restrictions on the other covariates, and to have constrained the coefficients of maternal age would have significantly reduced model fit.

The ordered birth weight data provided opportunity to consider the choice of model, investigate the merits of the fitted models, and probe the interpretation of the estimates in a real world setting. The modeling assumptions underlying the CR probability model were untenable for an outcome variable that arose from categorization of birth weight, because the birth weight categories did not truly arise as a hierarchy of binary events, but otherwise there was much to recommend this model. The constraint imposed in fitting the CR probability model produced far less loss of model fit than the constraints imposed in the other two models, and the summary estimate—a 24 percent reduction in the conditional probability of higher birth weight outcomes for the infants of mothers who smoked during pregnancy—was interpretable and plausible. Reasonably in the circumstances, it fell between the estimates from the AC and PP models.

The limitations of this work should be kept in mind. Firstly, the results are based on a simple data example for a dichotomous covariate, a single set of data simulations

for a continuous covariate and a response with just three ordered levels, and an analysis of a single real-world dataset in which issues with covariate scaling and statistical interaction were ignored. Secondly, results have been reported for the forwards-descending version of each model without investigation of the merits of the alternative versions. Thirdly, loss of model fit due to the imposition of constraints on the coefficients was investigated, but not the overall fit of the models. Fourthly, the variance estimates are taken from the inverse of the observed information matrix without thorough investigation of using the expected information matrix in its place, or of using an "information sandwich" to provide robustness to model misspecification. The non-robust estimator based on observed information provided confidence interval coverage for the ordinal response models that was about right, however. Fifthly, we used results based on maximum likelihood theory to compute Wald-type confidence intervals when the functional form of the log-link models does not satisfy theoretical assumptions about the parameter space required for distributional theory. Sixthly, our investigation of power was limited to a single type of model misspecification. It ignored other forms of departure such as omitted covariates and an incorrect link function, and other issues including sample size and model complexity, that would be taken into account in a more complete investigation. Finally, we omitted a thorough comparison of the merits of "partially-constrained" and "fully-constrained" models, and of models with non-linear constraints [21]. These may be important in practice as a means of improving model fit when a linear constraint applied across some covariates is untenable.

9. Conclusion

The three ordinal response log-link models offer a practical solution to the problem of obtaining adjusted estimates of the risk ratio for ordinal outcome data. Particularly when the model-based probabilities are high, however, the models must be used with some care and attention to detail.

10. Acknowledgements

REFERENCES

[1]   L. A. Goodman. "The Analysis of Dependence in Cross-Classifications Having Ordered Categories, Using Log-Linear Models for Frequencies and Log-Linear Models for Odds," *Biometrics*, Vol. 39, No. 1, 1983, pp. 149-160. doi:10.2307/2530815

[2]   S. E. Fienberg, "The Analysis of Cross-Classified Data,"

2nd Edition, MIT Press, Cambridge, 1980.

[3]   S. H. Walker and D. B. Duncan, "Estimation of the Probability of an Event as a Function of Several Independent Variables," *Biometrika*, Vol. 54, No. 1, 1967, pp. 167-79.

[4]   P. McCullough, "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society* B, Vol. 42, No. 2, 1980, pp. 109-142.

[5]   D. McFadden, "Econometric Models for Probabilistic Choice among Products," *Journal of Business*, Vol. 53, No. 3, 1980, pp. S13-S29. doi:10.1086/296093

[6]   D. W. Hosmer and S. Lemeshow, "Applied Logistic Regression," 2nd Edition, Wiley, New York, 2000. doi:10.1002/0471722146

[7]   L. Blizzard and D. W. Hosmer, "Parameter Estimation and Goodness-of-Fit in Log Binomial Regression," *Biometrical Journal*, Vol. 48, No. 6, 2006, pp. 5-22.

[8]   O. S. Miettenen, "Theoretical Epidemiology," Wiley, New York, 1985.

[9]   S. Greenland, "Interpretation and Choice of Effect Measures in Epidemiologic Analysis," *American Journal of Epidemiology*, Vol. 125, No. 5, 1987, pp. 761-768.

[10] O. S. Miettenen and E. F. Cook, "Confounding: essence and detection," *American Journal of Epidemiology*, Vol. 114, No. 4, 1981, pp. 593-603.

[11] J. Lee, "Odds Ratio or Relative Risk for Cross-Sectional Data?" *International Journal of Epidemiology*, Vol. 23, No. 1, 1994, pp. 201-203. doi:10.1093/ije/23.1.201

[12] C. C. Clogg and E. S. Shihadeh, "Statistical Models for Ordinal Variables," Sage, Thousand Oaks, 1994.

[13] R. Bender and A. Benner, "Calculating Ordinal Regression Models in SAS and S-Plus," *Biometrical Journal*, Vol. 42, No. 6, 2000, pp. 677-699. doi:10.1002/1521-4036(200010)42:6<677::AID-BIMJ677>3.0.CO;2-O

[14] A. A. O'Connell, "Logistic Regression Models for Ordinal Response Variables," Sage, Thousand Oaks, 2006.

[15] B. G. Armstrong and M. Sloan, "Ordinal Regression Models for Epidemiologic Data," *American Journal of Epidemiology*, Vol. 129, No. 1, 1989, pp. 191-204.

[16] D. Katz, J. Baptista, S. P. Azen and M. C. Pike, "Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies," *Biometrics*, Vol. 34, 1978, pp. 469-474. doi:10.2307/2530610.

[17] C. R. Rao, "Linear Statistical Inference and Its Applications," 2nd Edition, Wiley, New York, 2002.

[18] J. Neyman and E. S. Pearson "On the Use and Interpretation of Certain Test Criteria," *Biometrika*, Vol. 20, No. 1/2, 3/4, 1928, pp. 175-240, 263-294.

[19] C. R. Rao, "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation," *Proceedings of the Cambridge Philosophical Society*, Vol. 44, No. 1, 1948, pp. 50-57. doi:10.1017/S0305004100023987.

[20] A. Wald, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical So-*

ciety, Vol. 54, No. 3, 1943, pp. 426-482. doi:10.1090/S0002-9947-1943-0012401-3.

[21] B. Peterson and F. E. Harrrell, "Partial Proportional Odds Models for Ordinal Response Variables," *Applied Statis-*

tics, Vol. 39, No. 2, 1990, pp. 2005-2017. doi:10.2307/2347760.

[22] P. McCullough and J. Nelder, "Generalized Linear Models," 2nd Edition, Chapman Hall, New York, 1989.