



Bias correction for the proportional odds logistic regression model with application to a study of surgical complications

Stuart R. Lipsitz,

Brigham and Women's Hospital, Boston, USA

Garrett M. Fitzmaurice,

Harvard Medical School, Boston, USA

Scott E. Regenbogen,

University of Michigan, Ann Arbor, USA

Debajyoti Sinha,

Florida State University, Tallahassee, USA

Joseph G. Ibrahim

University of North Carolina at Chapel Hill, USA

and Atul A. Gawande

Brigham and Women's Hospital, Boston, USA

[Received October 2010. Final revision May 2012]

Summary. The proportional odds logistic regression model is widely used for relating an ordinal outcome to a set of covariates. When the number of outcome categories is relatively large, the sample size is relatively small and/or certain outcome categories are rare, maximum likelihood can yield biased estimates of the regression parameters. Firth and Kosmidis proposed a procedure to remove the leading term in the asymptotic bias of the maximum likelihood estimator. Their approach is most easily implemented for univariate outcomes. We derive a bias correction that exploits the proportionality between Poisson and multinomial likelihoods for multinomial regression models. Specifically, we describe a bias correction for the proportional odds logistic regression model, based on the likelihood from a collection of independent Poisson random variables whose means are constrained to sum to 1, that is straightforward to implement. The method proposed is motivated by a study of predictors of post-operative complications in patients undergoing colon or rectal surgery.

Keywords: Discrete response; Multinomial likelihood; Multinomial logistic regression; Penalized likelihood; Poisson likelihood

1. Introduction

Categorical responses which are ordinal in nature commonly arise in studies in the health,

Address for correspondence: Stuart R. Lipsitz, Division of General Internal Medicine, Brigham and Women's Hospital, 3rd Floor, BC3 002D, 1620 Tremont Street, Boston, MA 02120-1613, USA.
E-mail: slipsitz@partners.org

behavioural and social sciences. For example, in an epidemiological study of behavioural risk factors for stroke, the severity of a patient's stroke may be defined on an ordinal scale categorized as minor, moderate or severe. The proportional odds logistic regression model is probably the most widely used model for relating an ordinal outcome to a set of covariates. Typically, maximum likelihood (ML) is the method of choice for estimating the regression parameters. However, when the number of outcome categories is relatively large, the sample size is relatively small and/or some of the outcome categories are rare, ML can yield biased estimates of the regression parameters. Firth (1993) and Kosmidis and Firth (2009) proposed a procedure to remove the leading term in the asymptotic bias of the ML estimator. This approach is most easily implemented for univariate outcomes, e.g. Bernoulli and Poisson outcomes. The focus of this paper is on bias-corrected estimates of the regression parameters of the proportional odds logistic regression model.

For the multinomial logistic regression model for nominal (unordered) responses, Bull *et al.* (2002) proposed a penalized likelihood approach to remove the first-order bias of the ML estimate; their bias reducing score functions involve Kronecker product matrix operators and matrices of third-order derivatives. Recently, Kosmidis and Firth (2011) exploited the connection between multinomial logistic regression models and Poisson log-linear models for cell counts (Birch, 1963) to produce an approach to bias correction based on univariate Poisson likelihoods. This elegant approach requires the addition of nuisance parameters to the Poisson log-linear model that correspond to the multinomial totals for each subject. We note that, even though the implementation is different, the approaches of Bull *et al.* (2002) and Kosmidis and Firth (2011) are both based on a penalized multinomial likelihood and thus lead to the same bias-corrected estimates.

The approach of Kosmidis and Firth (2011), however, is restricted to multinomial models that can be expressed as log-linear models, i.e. multinomial logistic regression (McCullagh and Nelder, 1989). Importantly, this excludes applications of the method to the proportional odds logistic regression model, the non-proportional odds model or indeed any multinomial model with a non-canonical link function (e.g. a probit or complementary log-log-link). Because the multinomial proportional odds model is considered a multivariate generalized linear model, it falls within the general class of multivariate models that were considered in Kosmidis and Firth (2009). Kosmidis and Firth (2009) derived general expressions for the adjusted score equations for these multivariate models, and these adjusted score equations can be used to formulate a bias-corrected estimate for the proportional odds model. Instead of using these general adjusted score equations for multinomial regression models, here, we propose to obtain the bias-corrected estimates for the proportional odds model via iterative updates of pseudoresponses for univariate Poisson likelihoods.

In particular, for any multinomial regression model in which the probabilities are formulated to sum to 1 (such as the proportional odds model), we show that the multinomial likelihood is proportional to the likelihood from a collection of independent Poisson random variables. Thus, although the proportional odds logistic regression model cannot be expressed as a log-linear model so the method of Kosmidis and Firth (2011) does not apply, we can use a Poisson likelihood to solve the bias-corrected score equations in terms of simple iterative updates of pseudoresponses for univariate Poisson likelihoods, as opposed to using the general formulation in Kosmidis and Firth (2009) for multinomial likelihoods. For example, using our approach with the proposed pseudoresponses, it is relatively straightforward to implement the bias correction within existing statistical software (e.g. SAS procedure NLMIXED). Thus, the potential advantage of our proposed method is in terms of ease of implementation.

We emphasize that, even though both our approach for the proportional odds model and

Kosmidis and Firth's (2011) approach for multinomial logistic regression models use a likelihood that is a product of univariate Poisson distributions, our approach is for multinomial regression models in which the probabilities (Poisson means), by construction, are formulated to sum to 1, whereas Kosmidis and Firth's (2011) approach relies on Poisson log-linear models in which the Poisson means are more generally formulated to be expected counts. We discuss these differences in Appendix B of this paper.

The method proposed is motivated by a study of predictors of post-operative complications (Gawande *et al.*, 2007). In this study, 102 patients undergoing colon or rectal surgery at Brigham and Women's Hospital in Boston, Massachusetts, USA, were evaluated for predictors of the ordinal outcome 'major post-operative complications' (1, none; 2, major complication; 3, death) within 30 days post surgery, as part of the hospital's 'National surgical quality improvement program' cohort. In this programme, a systematic sample of patients undergoing general or vascular surgery in participating institutions were evaluated by trained, audited surgical research nurses for preoperative comorbidities and post-operative events within 30 days of surgery. The main predictor of interest in this study is the so-called 'surgical Apgar score', which is a 10-point measure that gauges intraoperative safety, according to blood loss, lowest heart rate and lowest mean arterial pressure obtained during the operation. A score of 0 denotes a poor prognosis, whereas a score of 10 is the best prognosis for recovery without complications. In previous analyses, Gawande *et al.* (2007) categorized the surgical Apgar score into five categories: scores ranging from 0–2, 3–4, 5–6, 7–8 and 9–10; further, in their analyses they treated these categories as nominal, not ordinal. A second predictor of interest is the American Society of Anesthesiologists score, which is a global assessment of the physical status of the patient before surgery (Owens *et al.*, 1978). This score yielded a binary indicator of preoperative disease status (1, systemic or worse disease; 0, mild or no disease). The question of scientific interest in this study was whether the preoperative disease status and intra-operative surgical Apgar score can predict patients who will have post-operative complications; the ability to discriminate patients in this way would allow surgeons to alter the amount and intensity of post-operative monitoring and care appropriately. Table 1 presents descriptive statistics and the results of separate bivariate analyses of the associations between the ordinal post-operative complications outcome and these two predictors. Note, in this sample, that there are no patients with surgical Apgar scores in the 0–2-range. The test for association between post-operative complications and the surgical Apgar score is based on a Kruskal–Wallis exact test (treating the categorical surgical Apgar

Table 1. Bivariate analyses for the post-operative surgical complications data (counts and row percentages presented)

Variable	Level	Post-operative complications			P-value
		None	Complication	Death	
Overall		83 (81%)	15 (15%)	4 (4%)	
Surgical Apgar score	3–4	3 (37.5%)	3 (37.5%)	2 (25.0%)	0.003†
	5–6	19 (76.0%)	5 (20.0%)	1 (4.0%)	
	7–8	51 (87.9%)	6 (10.3%)	1 (1.7%)	
	9–10	10 (90.9%)	1 (9.1%)	0 (0.0%)	
Preoperative disease	No	52 (83.9%)	8 (12.9%)	2 (3.2%)	0.425‡
	Yes	31 (77.5%)	7 (17.5%)	2 (5.0%)	

†Exact *P*-value for a Kruskal–Wallis test (treating surgical Apgar score as nominal).

‡Exact *P*-value for a Wilcoxon rank sum test.

predictor as nominal). The test for association between post-operative complications and the binary American Society of Anesthesiologists score is based on an exact Wilcoxon test. The preliminary results in Table 1 indicate that the surgical Apgar score is significantly associated with post-operative complications, but preoperative disease status is not.

In the medical literature, models for complications often differ by gender; for example, gender differences have been found in cardiac surgery (Guru *et al.*, 2006), thoracic surgery (Falcoz *et al.*, 2007) and vascular surgery (Nguyen *et al.*, 2009). Thus, it is of secondary interest to examine the associations separately for males and females. Because it is of interest to examine the joint effects of disease status and surgical Apgar score on post-operative complications, we initially fit a cumulative logistic model (setting ‘no complications’ as the reference category for the ordinal outcome) with surgical Apgar and American Society of Anesthesiologists scores as nominal and dichotomous covariates respectively. For the overall sample (102 patients), ML estimates of the regression parameters produced by three widely used software packages (SAS PROC LOGISTIC, the R function `polr` and the Stata command `ologit`) were identical. However, in analyses that were restricted to the males sample, none of these three packages converged to a unique solution to the ML equations. Further, with a total of only 19 complications or deaths overall (10 in males and nine in females), the standard ML estimates could potentially be badly biased. These observations led us to explore alternative approaches that yield less biased estimates of the proportional odds logistic regression parameters in small samples.

In Section 2, we briefly describe the underlying multinomial distribution for the categorical response, and we show that the corresponding likelihood can be expressed as the likelihood from a collection of independent Poisson random variables whose means are constrained to sum to 1. We describe a general bias correction for this Poisson formulation of the likelihood. In Section 3, we apply the bias correction to the proportional odds logistic regression model. In Section 4, we apply this approach in regression analyses of the data from the study of post-operative complications (Gawande *et al.*, 2007). In Section 5, we present results of a small-scale simulation study of bias correction for the proportional odds model. In the example and simulations, we also compare our approach with the *ad hoc* bias reduction approach that was proposed by Clogg *et al.* (1991); the latter approach adds a small constant to each subject’s multinomial outcome in the sample.

2. Multinomial and Poisson likelihoods for categorical data

Suppose that we have n independent subjects, where the i th individual’s ($i = 1, 2, \dots, n$) response Y_i is multinomial and, without loss of generality, can equal any value in ($j = 1, \dots, J$). We let the indicator random variable Y_{ij} equal 1 if the i th individual has response value j and equal 0 otherwise, with $\sum_{j=1}^J Y_{ij} = 1$. Each individual is assumed to have a $Q \times 1$ vector of covariates $\mathbf{x}_i = (x_{i1}, \dots, x_{iQ})'$. Note that we do not define \mathbf{x}_i here to include intercepts for separate multinomial levels; it contains only the subject covariates such as age and gender. Then, we denote the probability of response j given \mathbf{x}_i as

$$p_{ij} = p_{ij}(\beta) = \text{pr}(Y_i = j | \mathbf{x}_i, \beta) = \text{pr}(Y_{ij} = 1 | \mathbf{x}_i, \beta)$$

where β is a $R \times 1$ vector of parameters, and $\sum_{j=1}^J p_{ij} = 1$. In general, the vector β can contain different intercepts, and possibly different regression coefficients, for each multinomial level j (hence R , the dimension of β , is greater than Q , the dimension of the covariate vector \mathbf{x}_i). The model for p_{ij} for the proportional odds model (the focus of this paper) is given in the following section.

The probability mass function for subject i is multinomial

$$f(y_{i1}, y_{i2}, \dots, y_{iJ}) = \prod_{j=1}^J p_{ij}^{y_{ij}}.$$

Next, we show that the multinomial likelihood can be transformed into a Poisson likelihood as long as $\sum_{j=1}^J p_{ij} = 1$. If the Y_{ij} were independent Poisson random variables with mean $E(Y_{ij}) = p_{ij}$ then the corresponding Poisson likelihood would be proportional to

$$\exp(-p_{i+}) \prod_{j=1}^J p_{ij}^{y_{ij}}, \tag{1}$$

where $p_{i+} = \sum_{j=1}^J p_{ij}$. In this Poisson likelihood formulation, p_{i+} is required to be positive. If, however, the p_{ij} s are formulated so that they sum to 1 over the j s for every subject, then the Poisson and multinomial likelihoods are proportional. In Appendix A we show that the score equations for β and the expected information are identical under the multinomial and Poisson likelihood formulations (subject to the constraint that $\sum_{j=1}^J p_{ij} = 1$). Consequently, likelihood inferences will be identical on the basis of either likelihood. In particular, provided that the term $\exp(-p_{i+})$ is not a function of any unknown regression parameters (a condition that is satisfied if $p_{i+} = 1$ for all i) then the Poisson and multinomial likelihoods are proportional. In general, linear and log-linear models for p_{ij} are not formulated so that $p_{i+} = 1$, and thus the Poisson and multinomial likelihoods will not be proportional for these types of models.

Because $\sum_{j=1}^J p_{ij} = 1$ is satisfied for the proportional odds regression model (which is discussed in more detail in Section 3), removal of the first-order bias of the ML estimator for this model can be based on either the multinomial or Poisson likelihoods. The bias corrections yield identical results because the likelihoods are proportional, and the correction is based on the asymptotic variance of the parameter estimates which is shown in Appendix A to be the same under the two models. However, by substituting a Poisson likelihood (with constrained means) for the multinomial likelihood, it is more straightforward to base the bias correction on a likelihood that is formulated in terms of the product of univariate Poisson random variables than on a likelihood for a multinomial random variable. Next, we describe the score equations for the Poisson likelihood and discuss how the bias correction can be made in terms of iterative updates of ‘pseudoresponses’, which is an approach that was first described by Firth (1993). In Appendix A, we show that, from first principles without directly using the proportionality property, the score equations for the multinomial likelihood are identical to those for the Poisson likelihood.

The Poisson likelihood score equations for β are given by

$$u(\hat{\beta}) = \sum_{i=1}^n \sum_{j=1}^J \hat{D}_{ij} \hat{p}_{ij}^{-1} (y_{ij} - \hat{p}_{ij}) = 0, \tag{2}$$

where $D_{ij} = \partial p_{ij}(\beta) / \partial \beta$. Using the first-order bias correction for a univariate outcome given in Firth (1993) and Box (1971), the score equations (2) can be modified by replacing y_{ij} with the ‘pseudoresponse’

$$y_{ij}^* = y_{ij} + a_{ij}, \tag{3}$$

where

$$a_{ij} = 0.5 \text{tr}\{\text{var}(\hat{\beta}) D_{ij}^2\}. \tag{4}$$

In equation (4), $\text{var}(\hat{\beta})$ is the asymptotic variance–covariance matrix of $\hat{\beta}$ (estimated via the inverse of the observed or expected information matrix) and $D_{ij}^2 = \partial^2 p_{ij} / \partial \beta \partial \beta'$ is an $R \times R$ matrix of second derivatives of p_{ij} with respect to the $R \times 1$ vector of parameters β . To obtain the first-order bias-corrected estimate of β , one can iterate between updating y_{ij}^* given a current estimate of β , and then re-estimating β given the updated y_{ij}^* by solving equations (2), until the estimates of β converge. Typically, the quantities a_{ij} in equation (3) are small and positive, and their inclusion tends to reduce the effect of sampling 0s (so-called ‘empty cells’ or ‘0-cells’) and to increase the likelihood of convergence; however, in general, there is no guarantee of convergence. Finally, it is worth re-emphasizing that this bias correction based on a Poisson likelihood formulation does require a model for p_{ij} that constrains $\sum_{j=1}^J p_{ij} = 1$. Next, we consider a specific application of this bias correction to the proportional odds logistic regression model. The focus of this paper is on the proportional odds logistic regression model for ordinal responses; however, in Appendix B, we briefly discuss the implementation of our Poisson likelihood approach applied to a multinomial logistic regression model for nominal (unordered) responses, and we contrast our approach with the Poisson log-linear approach of Kosmidis and Firth (2011).

3. Proportional odds logistic regression model

The proportional odds logistic regression model can be written as

$$\pi_{ij} = \text{pr}(Y_i \leq j | \mathbf{x}_i, \beta) = \frac{\exp(\beta_{0j} + \beta_1' \mathbf{x}_i)}{1 + \exp(\beta_{0j} + \beta_1' \mathbf{x}_i)}, \tag{5}$$

for $j = 1, \dots, J - 1$ where π_{ij} is a ‘cumulative probability’, \mathbf{x}_i is a $Q \times 1$ vector of covariates as discussed above, β_{0j} is an intercept for cut point j and β_1 is a $Q \times 1$ vector of parameters. The regression parameters can be grouped together to form the $(Q + J - 1) \times 1$ vector $\beta = (\beta_{01}, \dots, \beta_{0,J-1}, \beta_1')'$. The probability of response level J , p_{iJ} , is

$$p_{iJ} = 1 - \pi_{i,J-1} = \frac{1}{1 + \exp(\beta_{0,J-1} + \beta_1' \mathbf{x}_i)}.$$

Then,

$$\begin{aligned} p_{ij} &= \text{pr}(Y_i = j | \mathbf{x}_i, \beta) \\ &= \text{pr}(Y_i \leq j | \mathbf{x}_i, \beta) - \text{pr}(Y_i \leq j - 1 | \mathbf{x}_i, \beta) \\ &= \pi_{ij} - \pi_{i,j-1}, \end{aligned} \tag{6}$$

for $j = 1, \dots, J$ where we define $\pi_{iJ} = 1$ and $\pi_{i0} = 0$ since $p_{iJ} = 1 - \pi_{i,J-1}$ and $p_{i1} = \pi_{i1} - 0$. The contribution to the likelihood for subject i can be rewritten as

$$\prod_{j=1}^J p_{ij}^{y_{ij}} = \prod_{j=1}^J (\pi_{ij} - \pi_{i,j-1})^{y_{ij}}.$$

The Poisson formulation of the multinomial likelihood is now used to obtain a simple bias correction term. As was discussed in Section 2, if we specify the Y_{ijs} as Poisson, the Poisson and multinomial likelihoods are proportional provided that $p_{i+} = 1$ for all subjects. Thus, we must show that $p_{i+} = 1$. Using the model for p_{ij} in equation (6),

$$\begin{aligned} \sum_{j=1}^J p_{ij} &= \sum_{j=1}^J (\pi_{ij} - \pi_{i,j-1}) \\ &= \sum_{j=1}^J \pi_{ij} - \sum_{j=0}^{J-1} \pi_{ij} \\ &= \pi_{iJ} - \pi_{i0} \\ &= 1, \end{aligned}$$

for all i . This establishes that the Poisson and multinomial likelihoods are proportional for the proportional odds logistic regression model.

To formulate the pseudoresponses in equation (3) that are required for the bias correction, we need to calculate D_{ij}^2 . For simplicity, we rewrite equation (5) as

$$\pi_{ij} = \frac{\exp(\beta' z_{ij})}{1 + \exp(\beta' z_{ij})},$$

where z_{ij} is a $(Q + J - 1) \times 1$ design vector that includes the $Q \times 1$ vector of covariates \mathbf{x}_i and also indicators for the intercepts at each response level, and $\beta = (\beta_{01}, \dots, \beta_{0,J-1}, \beta_1')'$ is described above. Then, for the cumulative logistic model,

$$D_{ij}^2 = \frac{\partial^2(\pi_{ij} - \pi_{i,j-1})}{\partial\beta \partial\beta'} = z_{ij} z'_{ij} \pi_{ij} (1 - \pi_{ij})(1 - 2\pi_{ij}) - z_{i,j-1} z'_{i,j-1} \pi_{i,j-1} (1 - \pi_{i,j-1})(1 - 2\pi_{i,j-1}),$$

and, for the pseudoresponse,

$$\begin{aligned} a_{ij} &= 0.5 \text{tr}\{\text{var}(\hat{\beta}) D_{ij}^2\} \\ &= 0.5 \text{tr}\{\text{var}(\hat{\beta}) \{z_{ij} z'_{ij} \pi_{ij} (1 - \pi_{ij})(1 - 2\pi_{ij}) - z_{i,j-1} z'_{i,j-1} \pi_{i,j-1} (1 - \pi_{i,j-1})(1 - 2\pi_{i,j-1})\}\} \\ &= 0.5 [\text{var}\{\text{logit}(\hat{\pi}_{ij})\} \pi_{ij} (1 - \pi_{ij})(1 - 2\pi_{ij}) - \text{var}\{\text{logit}(\hat{\pi}_{i,j-1})\} \pi_{i,j-1} (1 - \pi_{i,j-1})(1 - 2\pi_{i,j-1})]. \end{aligned}$$

Thus, for the Poisson formulation of the cumulative logistic regression model, the pseudo-response equals

$$\begin{aligned} y_{ij}^* &= y_{ij} + 0.5 [\text{var}\{\text{logit}(\hat{\pi}_{ij})\} \pi_{ij} (1 - \pi_{ij})(1 - 2\pi_{ij}) \\ &\quad - \text{var}\{\text{logit}(\hat{\pi}_{i,j-1})\} \pi_{i,j-1} (1 - \pi_{i,j-1})(1 - 2\pi_{i,j-1})], \end{aligned}$$

for $j = 2, \dots, J - 1$. Note, when $j = 1$, $p_{i1} = \pi_{i1}$, so

$$y_{i1}^* = y_{i1} + 0.5 \text{var}\{\text{logit}(\hat{\pi}_{i1})\} \pi_{i1} (1 - \pi_{i1})(1 - 2\pi_{i1})$$

and, when $j = J$, $p_{iJ} = 1 - \pi_{i,J-1}$, so

$$y_{iJ}^* = y_{iJ} - 0.5 \text{var}\{\text{logit}(\hat{\pi}_{i,J-1})\} \pi_{i,J-1} (1 - \pi_{i,J-1})(1 - 2\pi_{i,J-1}).$$

In principle, this bias-corrected approach based on a Poisson likelihood can be fitted within the generalized linear models framework. In practice, though, the mean for Y_{ij} under a proportional odds logistic model, $p_{ij} = E[Y_{ij} | z_{ij}, z_{i,j-1}] = \pi_{ij} - \pi_{i,j-1}$, is not a standard option for Poisson regression in widely used software for generalized linear models. Thus, we implemented the bias correction by using an optimization procedure for non-linear regression models, SAS PROC NL MIXED (SAS Institute, 2010). We note that the pseudoresponse is straightforward to calculate since the predicted cumulative probabilities $\hat{\pi}_{ij}$ and the variance of the predicted cumulative log-odds, $\text{var}\{\text{logit}(\hat{\pi}_{i,j-1})\}$, are typically standard output of any non-linear regression program, including SAS PROC NL MIXED. Thus, although the algorithm is not very complicated, it did require us to write a special purpose program to maximize the cumulative logistic

model via a Poisson likelihood. Specifically, an SAS macro, which embeds SAS PROC NLMIXED (SAS Institute, 2010), was written to implement the bias-corrected proportional odds regression estimator; SAS PROC NLMIXED calculates variances based on the inverse of the observed information. The SAS macro can be obtained from

<http://www.blackwellpublishing.com/rss>

4. Application to study of surgical complications

In this section, we apply the proposed methodology to the analysis of the surgical complications data that were described in Section 1. The study includes 102 patients undergoing colorectal surgery at Brigham and Women’s Hospital. The outcome is the ordinal variable ‘major post-operative complications’ (1, none; 2, major complication; 3, death) within 30 days post surgery. There are two main predictors of interest: the four-level categorical surgical Apgar score and the dichotomous preoperative disease status (1, systemic or worse disease; 0, mild or no disease) of the patient. *A priori*, our surgical colleagues conjectured that patients with worse (lower) surgical Apgar scores and systemic or worse preoperative disease would be more likely to have post-operative complications.

To examine the joint relationship between post-operative complications and these two covariates, we fit the proportional odds logistic regression model

$$\text{logit}(\pi_{ij}) = \text{logit}\{\text{pr}(Y_i \geq j | \mathbf{x}_i, \beta)\} = \beta_{0j} + \beta_1 \text{Apgar}(3:4)_i + \beta_2 \text{Apgar}(5:6)_i + \beta_3 \text{Apgar}(7:8)_i + \beta_4 \text{Disease}_i, \tag{7}$$

for $j = 2, 3$, where $\text{Apgar}(k:l)_i = 1$ denotes that the surgical Apgar score equals k or l (surgical Apgar score 9–10 is the reference category), and $\text{Disease}_i = 1$ if the patient has systemic or worse preoperative disease, and 0 otherwise. Note, in a slight departure from notation used in earlier sections where we defined π_{ij} in terms of cumulating over *lower* values of the ordinal outcome, for ease of interpretation, here we accumulate over *higher* values of the ordinal post-operative complications outcome. In particular, we model two ‘cumulative’ probabilities: the probability of complications or death $\text{pr}(Y_i \geq 2)$ and the probability of death $\text{pr}(Y_i = 3)$.

Table 2 gives the estimates of β obtained by using the bias-corrected method for the data based on the total sample ($n = 102$), as well as the standard ML estimates of β (the latter were obtained by using SAS PROC LOGISTIC). For comparison, we also give the results by using the *ad hoc* bias correction approach proposed by Clogg *et al.* (1991). The approach of Clogg *et al.* (1991) requires creation of $J - 1$ additional responses for each subject associated with the same covariates; these $J - 1$ additional observations take the $J - 1$ values that the original Y_i did not; $k = 1, \dots, J, k \neq Y_i$. The original Y_i is assigned weight $1 + 1/(nJ)$ and the $J - 1$ new observations are assigned weight $1/(nJ)$ in subsequent analysis that treats all observations as independent. This procedure effectively adds $nJ \times 1/(nJ) = 1$ observations to the original data set.

Of note, in Table 2, there were no convergence problems with ML for the analysis of data using the entire sample. However, there were some differences in the odds ratio estimates for the effects of surgical Apgar score obtained from the two approaches. For example, the estimated odds ratio for surgical Apgar score 3–4 *versus* surgical Apgar score 9–10 is $\exp(2.869) = 17.6$ by using standard ML and $\exp(2.440) = 11.5$ by using the bias-corrected estimator: a relative difference of 54%. The estimate from the approach of Clogg *et al.* (1991) fell in between, $\exp(2.785) = 16.2$, although closer to standard ML. When these estimates are compared with their standard errors, all three methods lead to the same conclusion that the largest effect on post-operative complications is for surgical Apgar score 3–4 *versus* surgical Apgar score 9–10

Table 2. Comparison of proportional odds logistic regression parameter estimates for the post-operative surgical complications data, full sample ($n = 102$)

<i>Effect</i>	<i>Approach</i>	<i>Estimate</i>	<i>Standard error</i>	<i>Z-statistic</i>	<i>P-value</i>
Intercept ($j \geq 2$)	Standard ML†	2.436	1.072	2.27	0.023
	ML: bias corrected	2.055	0.922	2.23	0.027
	Clogg <i>et al.</i> (1991)	2.363	1.038	2.28	0.023
Intercept ($j = 3$)	Standard ML†	4.385	1.186	3.70	< 0.001
	ML: bias corrected	3.819	1.026	3.72	< 0.001
	Clogg <i>et al.</i> (1991)	4.254	1.145	3.71	< 0.001
Apgar 3–4	Standard ML†	2.869	1.262	2.27	0.023
	ML: bias corrected	2.440	1.147	2.13	0.034
	Clogg <i>et al.</i> (1991)	2.785	1.230	2.26	0.024
Apgar 5–6	Standard ML†	1.156	1.155	1.00	0.317
	ML: bias corrected	0.845	1.015	0.83	0.406
	Clogg <i>et al.</i> (1991)	1.110	1.122	0.99	0.323
Apgar 7–8	Standard ML†	0.261	1.134	0.23	0.818
	ML: bias corrected	-0.029	0.992	-0.03	0.977
	Clogg <i>et al.</i> (1991)	0.246	1.099	0.22	0.823
Preoperative disease	Standard ML†	0.390	0.550	0.71	0.478
	ML: bias corrected	0.376	0.534	0.70	0.482
	Clogg <i>et al.</i> (1991)	0.376	0.541	0.69	0.487

†The standard ML approach is not bias corrected, with the convergence criterion: the relative change in the log-likelihood between successive iterations is less than 0.000001.

($P < 0.05$); however, the relative magnitudes of the effect estimates are discernibly different for the three methods. From the results of the three methods, the preoperative disease status does not appear to affect complications significantly in this sample.

As discussed in Section 1, predictive models for complications are often different for males ($n = 53$) and females ($n = 58$) (Guru *et al.*, 2006; Falcoz *et al.*, 2007; Nguyen *et al.*, 2009). Therefore, in secondary analyses, we examined the estimated effects for equation (7) separately for males and females. Table 3 presents descriptive statistics stratified by gender; the estimates of the regression parameters are given in Table 4. Although it is not immediately transparent from Table 3, when restricted to the sample of males, there is quasi-complete separation of data points. The definition of separation for ordinal data relies on the same definition as for binary data. For binary data, separation occurs when there is no overlap in the covariate values with $Y = 0$ and with $Y = 1$. Agresti (2010) defined separation for cumulative logit models (such as the proportional odds model) in terms of whether separation occurs for each of the possible collapsings of contiguous categories of the ordinal response to a binary response.

We note that the convergence criterion that is used for ML is that the relative change in the log-likelihood between successive iterations is less than 0.000001. The ML estimates for males that are reported in Table 4 are based on 10 iterations using the above convergence criterion (we note that three widely used software packages, SAS Proc LOGISTIC, Stata command `ologit` and the R function `polr`, all produced the same estimates with this convergence criterion). Although the likelihood converged to a finite value, many of the ML estimates in Table 4 for the sample of males appear to be diverging to ∞ . When there is quasi-complete (or complete) separation, the ML parameter estimates for the variable (or variables) with separation do not exist. In contrast, the bias-corrected estimator yields finite estimates that have been shown, in simulations (including in the following section), to have good sampling properties; however, we caution that somewhat greater care is required in interpreting the bias-corrected estimates

Table 3. Bivariate analyses for the post-operative surgical complications data, stratified by gender (counts and row percentages presented)

Variable	Level	Post-operative complications			P-value
		None	Complication	Death	
<i>Gender: male</i>					
Overall		43 (81.1%)	7 (13.2%)	3 (5.7%)	0.015†
Surgical Apgar score	3–4	1 (25.0%)	1 (25.0%)	2 (50.0%)	
	5–6	10 (83.3%)	1 (8.3%)	1 (8.3%)	
	7–8	29 (85.3%)	5 (14.7%)	0 (0.0%)	
	9–10	3 (100.0%)	0 (0.0%)	0 (0.0%)	
Preoperative disease	No	30 (85.7%)	3 (8.6%)	2 (5.7%)	0.354‡
	Yes	13 (72.2%)	4 (22.2%)	1 (5.6%)	
<i>Gender: female</i>					
Overall		49 (81.6%)	8 (16.3%)	1 (2.0%)	0.146†
Surgical Apgar score	3–4	2 (50.0%)	2 (50.0%)	0 (0.0%)	
	5–6	9 (69.2%)	4 (30.8%)	0 (0.0%)	
	7–8	22 (91.7%)	1 (4.2%)	1 (4.2%)	
	9–10	7 (87.5%)	1 (12.5%)	0 (0.0%)	
Preoperative disease	No	22 (81.5%)	5 (18.5%)	0 (0.0%)	0.999‡
	Yes	18 (81.8%)	3 (13.6%)	1 (4.6%)	

†Exact P-value for a Kruskal–Wallis test (treating the surgical Apgar score as nominal).

‡Exact P-value for a Wilcoxon rank sum test.

when there is quasi-complete separation. Also, the results of the approach of Clogg *et al.* (1991) for males give estimates that are much larger than the estimates from our bias-corrected procedure. On the basis of the bias-corrected analyses of the data (as opposed to standard ML or the method of Clogg *et al.* (1991)), the study investigators had greater confidence reporting the results for the total sample since the associations did not appear to differ by gender.

In summary, the results of analyses of the surgical complications data highlight how standard proportional odds logistic regression and the bias-corrected method can produce discernibly different estimates of effects. However, to examine the finite sample bias of these approaches, we conducted a simulation study; the results of the simulation study are reported in the next section.

5. Simulations for proportional odds model

In this section, we study the finite sample bias in estimating β for the proportional odds logistic regression model using ML, the bias-corrected method proposed in this paper, as well as the alternative approach for bias correction for multinomial regression models that was proposed by Clogg *et al.* (1991). We note that we present the results of our bias-corrected approach based on the observed information. We also ran simulations using the expected information and there was very little difference between using either the observed or expected information with respect to the bias correction.

We consider a proportional odds logistic regression model with three covariates:

$$\text{logit}(\pi_{ij}) = \beta_0j + \beta_1x_{i1} + \beta_2x_{i2} + \beta_3x_{i3}$$

$j = 1, \dots, J - 1$, where $J = 5$. We performed three sets of simulations. In all simulations, the intercepts were set to $\beta_0j = \text{logit}(j/J)$ (for $j < 5$).

Table 4. Comparison of proportional odds logistic regression parameter estimates for the post-operative surgical complications data, stratified by gender

Effect	Approach	Estimate	Standard error	Z-statistic	P-value
<i>Gender: male (n = 53)</i>					
Intercept ($j \geq 2$)	Standard ML†	11.778	208.500	0.06	0.955
	ML: bias corrected	1.920	1.741	1.10	0.272
	Clogg <i>et al.</i> (1991)	5.474	8.922	0.61	0.540
Intercept ($j = 3$)	Standard ML†	13.562	208.500	0.07	0.948
	ML: bias corrected	3.376	1.817	1.86	0.065
	Clogg <i>et al.</i> (1991)	7.223	8.946	0.81	0.419
Apgar 3–4	Standard ML†	12.988	208.500	0.06	0.950
	ML: bias corrected	2.858	2.087	1.37	0.173
	Clogg <i>et al.</i> (1991)	6.663	8.990	0.74	0.459
Apgar 5–6	Standard ML†	10.170	208.500	0.05	0.961
	ML: bias corrected	0.509	1.890	0.27	0.788
	Clogg <i>et al.</i> (1991)	3.889	8.954	0.43	0.664
Apgar 7–8	Standard ML†	9.690	208.500	0.05	0.963
	ML: bias corrected	-0.018	1.852	-0.01	0.992
	Clogg <i>et al.</i> (1991)	3.420	8.946	0.38	0.702
Preoperative disease	Standard ML†	0.596	0.822	0.72	0.469
	ML: bias corrected	0.545	0.796	0.68	0.495
	Clogg <i>et al.</i> (1991)	0.582	0.813	0.72	0.474
<i>Gender: female (n = 58)</i>					
Intercept ($j \geq 2$)	Standard ML†	1.938	1.112	1.74	0.081
	ML: bias corrected	1.576	0.972	1.62	0.107
	Clogg <i>et al.</i> (1991)	1.905	1.094	1.74	0.082
Intercept ($j = 3$)	Standard ML†	4.448	1.486	2.99	0.003
	ML: bias corrected	3.584	1.195	3.00	0.003
	Clogg <i>et al.</i> (1991)	4.331	1.437	3.01	0.003
Apgar 3–4	Standard ML†	1.833	1.460	1.26	0.209
	ML: bias corrected	1.491	1.330	1.12	0.264
	Clogg <i>et al.</i> (1991)	1.797	1.444	1.24	0.213
Apgar 5–6	Standard ML†	1.105	1.234	0.90	0.371
	ML: bias corrected	0.808	1.103	0.73	0.465
	Clogg <i>et al.</i> (1991)	1.080	1.217	0.89	0.375
Apgar 7–8	Standard ML†	-0.397	1.298	-0.31	0.760
	ML: bias corrected	-0.514	1.149	-0.45	0.655
	Clogg <i>et al.</i> (1991)	-0.379	1.274	-0.30	0.766
Preoperative disease	Standard ML†	-0.057	0.788	-0.07	0.943
	ML: bias corrected	-0.023	0.737	-0.03	0.975
	Clogg <i>et al.</i> (1991)	-0.054	0.778	-0.07	0.945

†The standard ML approach is not bias corrected, with the convergence criterion: the relative change in the log-likelihood between successive iterations is less than 0.000001.

For the first set of simulations, we let $(\beta_1, \beta_2, \beta_3) = (-1, -1, -0.06)$ and specified covariate distributions which gave approximately equal probabilities across all response categories. In particular, in the first set of simulations, the covariates were simulated independently with $x_{i1} \sim \text{Bern}(0.05)$, $x_{i2} \sim N(0, 1)$ and $x_{i3} \sim N(0, 8)$. For this first set of simulations, the average marginal probabilities are

$$n^{-1} \sum_{i=1}^n (p_{i1}, p_{i2}, p_{i3}, p_{i4}, p_{i5}) = (0.25, 0.17, 0.16, 0.17, 0.25).$$

In the second set of simulations, we again let $(\beta_1, \beta_2, \beta_3) = (-1, -1, -0.06)$ but specified covariate distributions which produced small probabilities in all response categories except

$J = 5$. In particular, the covariates were again simulated independently with $x_{i1} \sim \text{Bern}(0.05)$, $x_{i2} \sim N(0, 1)$, but with x_{i3} distributed as log-normal with a median of 54 and scale parameter 0.35 (the latter distribution is similar to that for age of adults). For this second set of simulations, the average marginal response probabilities are

$$n^{-1} \sum_{i=1}^n (p_{i1}, p_{i2}, p_{i3}, p_{i4}, p_{i5}) = (0.01, 0.02, 0.03, 0.07, 0.87).$$

In the third set of simulations, to explore possible problems caused by a large regression parameter, we let $(\beta_1, \beta_2, \beta_3) = (4, 1, 0.06)$. The covariate distributions were the same as in the first set of simulations: $x_{i1} \sim \text{Bern}(0.05)$ and $x_{i2} \sim N(0, 1)$ and $x_{i3} \sim N(0, 8)$. However, with $(\beta_1, \beta_2, \beta_3) = (4, 1, 0.06)$, this configuration produced small probabilities in all response categories except $J = 1$. In particular, for this third set of simulations, the average marginal probabilities are

$$n^{-1} \sum_{i=1}^n (p_{i1}, p_{i2}, p_{i3}, p_{i4}, p_{i5}) = (0.91, 0.05, 0.02, 0.01, 0.01).$$

Owing to the small probabilities that are associated with the majority of response categories, we expect the second and third sets of simulations to produce larger biases for standard ML.

We conducted simulations for two different sample sizes: $n = 40$ and $n = 80$. For each simulation configuration, 2500 simulation replications were performed. The convergence criterion for ML is that the relative change in the log-likelihood between successive iterations is less than 0.000001; we report the percentage of simulation replications in which this convergence criterion was not met. When ML fails to converge, we use the estimates from the 25th iteration (the default maximum number of iterations in SAS Proc LOGISTIC).

Tables 5–7 present the relative biases defined as $100(\hat{\beta} - \beta)/\beta$, the root-mean-square error (RMSE) and the coverage probabilities of 95% Wald confidence intervals for the three sets of simulations. We present results for all simulation replications, and also for the subset of simu-

Table 5. Simulation results with intercepts $\beta_{0j} = \text{logit}(j/J)$, $J = 5$, $(\beta_1, \beta_2, \beta_3) = (-1, -1, -0.06)$ and average $(p_{i1}, p_{i2}, p_{i3}, p_{i4}, p_{i5}) = (0.25, 0.17, 0.16, 0.17, 0.25)^\dagger$

Parameter	Sample size	Method	% relative bias	RMSE	Coverage probability (%)
$\beta_1 = -1$	40	ML	10.6	0.727	94.0
		ML: bias corrected	1.5	0.636	96.0
		Clogg <i>et al.</i> (1991)	6.6	0.667	94.9
	80	ML	7.0	0.446	95.1
		ML: bias corrected	-1.4	0.434	95.4
		Clogg <i>et al.</i> (1991)	4.1	0.449	94.1
$\beta_2 = -1$	40	ML	11.2	0.393	94.6
		ML: bias corrected	0.4	0.392	96.3
		Clogg <i>et al.</i> (1991)	7.0	0.392	96.0
	80	ML	5.6	0.256	94.7
		ML: bias corrected	0.2	0.246	95.4
		Clogg <i>et al.</i> (1991)	4.9	0.298	94.5
$\beta_3 = -0.06$	40	ML	13.2	0.045	94.2
		ML: bias corrected	1.6	0.043	96.1
		Clogg <i>et al.</i> (1991)	7.5	0.049	94.3
	80	ML	6.9	0.027	94.0
		ML: bias corrected	-0.8	0.024	95.8
		Clogg <i>et al.</i> (1991)	3.4	0.028	95.1

† All simulation replications converged for the ML method.

lation replications when ML converges. The latter results can be considered ‘conditional on the likelihood convergence criterion’.

For the first set of simulations (Table 5), with approximately equal probabilities in all five categories, the standard ML estimate has relative bias greater than 10% for all parameters when $n = 40$, and between 5% and 10% for $n = 80$. In contrast, the bias-corrected approach has negligible bias for both samples sizes. The approach of Clogg *et al.* (1991) has between 5% and 10% relative bias for $n = 40$, and less than 5% for $n = 80$; in general, the approach of Clogg *et al.* (1991) has greater relative bias than the bias-corrected approach, but less than standard ML. Overall, the RMSE is slightly smaller for the bias-corrected approach *versus* both the approach

Table 6. Simulation results with intercepts $\beta_{0j} = \text{logit}(j/J)$, $J = 5$, $(\beta_1, \beta_2, \beta_3) = (-1, -1, -0.06)$ and average $(p_{i1}, p_{i2}, p_{i3}, p_{i4}, p_{i5}) = (0.01, 0.02, 0.03, 0.07, 0.87)$

Parameter	Samples size	Method	% relatives bias	RMSE	Coverage probability (%)
<i>All simulation replications†</i>					
$\beta_1 = -1$	40	ML	254.4	3.330	97.7
		ML: bias corrected	-4.9	1.001	97.1
		Clogg <i>et al.</i> (1991)	12.7	0.933	97.9
	80	ML	29.4	0.767	94.8
		ML: bias corrected	-1.8	0.728	96.1
		Clogg <i>et al.</i> (1991)	4.8	0.704	96.9
$\beta_2 = -1$	40	ML	62.1	2.882	95.1
		ML: bias corrected	-4.4	0.623	97.1
		Clogg <i>et al.</i> (1991)	5.8	0.639	96.3
	80	ML	18.7	0.453	93.5
		ML: bias corrected	-2.3	0.406	95.7
		Clogg <i>et al.</i> (1991)	4.2	0.395	95.3
$\beta_3 = -0.06$	40	ML	123.8	0.192	95.3
		ML: bias corrected	-2.6	0.070	97.0
		Clogg <i>et al.</i> (1991)	14.4	0.068	97.5
	80	ML	29.1	0.071	94.7
		ML: bias corrected	1.6	0.047	96.3
		Clogg <i>et al.</i> (1991)	2.7	0.056	95.7
<i>Results when ML converged‡</i>					
$\beta_1 = -1$	40	ML	24.4	1.399	96.4
		ML: bias corrected	-15.0	0.878	97.3
		Clogg <i>et al.</i> (1991)	8.1	0.846	97.2
	80	ML	16.6	0.749	96.5
		ML: bias corrected	-2.4	0.636	96.7
		Clogg <i>et al.</i> (1991)	3.8	0.694	95.6
$\beta_2 = -1$	40	ML	25.1	0.678	95.3
		ML: bias corrected	-0.2	0.647	95.7
		Clogg <i>et al.</i> (1991)	2.7	0.483	96.2
	80	ML	11.8	0.447	95.6
		ML: bias corrected	1.5	0.388	95.6
		Clogg <i>et al.</i> (1991)	4.0	0.346	95.9
$\beta_3 = -0.06$	40	ML	45.1	0.123	96.0
		ML: bias corrected	-2.5	0.093	97.5
		Clogg <i>et al.</i> (1991)	-1.3	0.077	97.0
	80	ML	16.9	0.061	94.6
		ML: bias corrected	-0.6	0.052	96.0
		Clogg <i>et al.</i> (1991)	2.6	0.054	96.2

†When the ML method did not converge, the ML estimates are from the last (25th) iteration.

‡The ML method converged for 90% of the simulation replications when $n = 40$ and 99% when $n = 80$.

of Clogg *et al.* (1991) and standard ML. With simulation standard errors for coverage probabilities of approximately 0.44%, the coverage probabilities for both sample sizes and across all approaches attain the nominal 95% level.

For the second set of simulations (Table 6), with small probabilities in the first four response categories, ML converged for 90% of the simulation replications when $n = 40$, and 99% when $n = 80$. From the results including all simulation replications, it is apparent that the relative bias of the ML estimate can be very large in small samples ($n = 40$), with relative bias as large as 250%. Applying the bias correction to ML proposed in this paper reduces the bias to minimal levels (less than 5%). With $n = 80$, the ML approach can still yield appreciable bias, whereas

Table 7. Simulation results with intercepts $\beta_{0j} = \text{logit}(j/J)$, $J = 5$, $(\beta_1, \beta_2, \beta_3) = (4, 1, 0.06)$ and average $(\rho_{i1}, \rho_{i2}, \rho_{i3}, \rho_{i4}, \rho_{i5}) = (0.91, 0.05, 0.02, 0.01, 0.01)$

Parameter	Sample size	Method	% relatives bias	RMSE	Coverage probability (%)
<i>All simulation replications†</i>					
$\beta_1 = 4$	40	ML	115.1	7.338	97.8
		ML: bias corrected	-4.7	1.011	96.1
		Clogg <i>et al.</i> (1991)	1.9	1.117	94.0
	80	ML	97.5	6.632	94.6
		ML: bias corrected	-3.1	0.889	94.4
		Clogg <i>et al.</i> (1991)	5.1	1.048	96.0
$\beta_2 = 1$	40	ML	33.9	1.005	95.6
		ML: bias corrected	0.7	0.424	96.3
		Clogg <i>et al.</i> (1991)	9.7	0.533	94.4
	80	ML	12.1	0.400	94.6
		ML: bias corrected	-0.2	0.326	95.9
		Clogg <i>et al.</i> (1991)	1.3	0.340	96.0
$\beta_3 = 0.06$	40	ML	37.6	0.098	94.5
		ML: bias corrected	2.4	0.067	96.5
		Clogg <i>et al.</i> (1991)	-7.5	0.058	95.2
	80	ML	8.3	0.080	94.6
		ML: bias corrected	1.1	0.043	95.9
		Clogg <i>et al.</i> (1991)	2.2	0.046	94.8
<i>Results when ML converged‡</i>					
$\beta_1 = 4$	40	ML	-2.4	1.204	96.2
		ML: bias corrected	-20.3	1.123	97.3
		Clogg <i>et al.</i> (1991)	-21.9	1.202	86.2
	80	ML	-5.5	0.705	95.6
		ML: bias corrected	-14.6	0.822	92.0
		Clogg <i>et al.</i> (1991)	-8.8	0.695	94.3
$\beta_2 = 1$	40	ML	26.1	0.763	95.4
		ML: bias corrected	2.9	0.403	97.2
		Clogg <i>et al.</i> (1991)	12.4	0.556	96.3
	80	ML	11.5	0.391	93.9
		ML: bias corrected	-0.3	0.318	95.4
		Clogg <i>et al.</i> (1991)	4.7	0.333	96.1
$\beta_3 = 0.06$	40	ML	34.6	0.089	94.8
		ML: bias corrected	0.4	0.077	97.3
		Clogg <i>et al.</i> (1991)	-19.6	0.051	96.3
	80	ML	8.9	0.078	94.1
		ML: bias corrected	2.9	0.042	96.2
		Clogg <i>et al.</i> (1991)	-1.0	0.045	93.2

†When the ML method did not converge, the ML estimates are from the last (25th) iteration.

‡The ML method converged for 57% of the simulation replications when $n = 40$ and 63% when $n = 80$.

applying the first-order correction to ML results in negligible bias. The approach of Clogg *et al.* (1991) gives much smaller bias than standard ML (between 10% and 15% for $n = 40$, and less than 5% when $n = 80$). The RMSE is similar for the bias-corrected approach and the approach of Clogg *et al.* (1991) but can be much larger for standard ML. Although some of the coverage probabilities are as high as 97%, in general, the coverage probabilities appear to agree with the nominal 95% level. When restricted to the simulation replications where ML converged, as might be expected, there is far less bias for standard ML when compared with the results from all simulation replications.

The third set of simulations (Table 7), with small probabilities in the last four response categories (and a large $\beta_1 = 4$), give similar results to those of the second set. ML converged much less often than for the second set, with 57% of the simulation replications converging when $n = 40$, and 63% when $n = 80$. From the results including all simulation replications, the relative bias of the ML estimate can be very large in small samples ($n = 40$), with relative bias as large as 115%. Applying the bias correction to ML that is proposed in this paper reduces the bias to minimal levels (again less than 5%). With $n = 80$, the ML approach can still yield appreciable bias, whereas applying the first-order correction to ML results in negligible bias. The approach of Clogg *et al.* (1991) gives much smaller bias than standard ML (between 5% and 10% for $n = 40$, and less than 5% when $n = 80$). In general, the RMSE is slightly smaller for the bias-corrected approach *versus* the approach of Clogg *et al.* (1991) and again can be much larger for standard ML. When restricted to the simulation replications where ML converged, again there is far less bias for standard ML when compared with the results from all simulation replications; similarly to before, the bias-corrected approach and the approach of Clogg *et al.* (1991) tend to have greater bias when compared with the results from all simulation replications.

Although Wald confidence intervals are known to be conservative (Hauck and Donner, 1977; Heinze and Schemper, 2002; Bull *et al.*, 2007) with large β s, we found in the last set of simulations with $\beta_1 = 4$ that the coverage probabilities agree with the nominal 95% level. However, we cannot generalize on the basis of this one simulation set-up, so one would still want alternatives to obtain confidence intervals. On the basis of the results of theorem 1 of Kosmidis and Firth (2009), we cannot use a penalized likelihood approach with the proportional odds model to obtain confidence intervals; thus we suggest using the bootstrap as an alternative to obtain confidence intervals with large estimated regression coefficients.

6. Conclusion

In this paper we have described a simple implementation of Firth's (1993) bias correction in the proportional odds logistic regression model. By exploiting the connection between the multinomial and Poisson likelihoods (subject to a model that constrains the means to sum to 1 within subjects), we derived a bias correction based on univariate Poisson distributions. This bias correction adds a function of both the 'predicted probabilities' and the 'variance of the linear predictor' to the indicator for each outcome category; this in turn is used to form a pseudoresponse that replaces the original indicator. This pseudoresponse is relatively simple to calculate and leads to an iterative algorithm that is straightforward to implement. Because the proportional odds model is probably the most widely used regression model for ordinal categorical data, the approach to bias correction that is described here should be useful to applied statisticians.

Although not specifically discussed in this paper, the method proposed can also be used for any multinomial model that constrains the multinomial probabilities to sum to 1, including the non-proportional odds model (Williams and Grizzle, 1972) and multinomial models with non-canonical link functions (e.g. probit or complementary log-log-link). We note that

Kosmidis and Firth’s (2011) bias correction approach was specifically developed for the multinomial logistic regression model for nominal (unordered) data. In particular, Kosmidis and Firth (2011) applied Birch’s (1963) connection between Poisson log-linear models for cell counts and multinomial logistic regression models; this requires the addition of a nuisance parameter to the Poisson log-linear model for each subject. This nuisance parameter corresponds to the multinomial total for each subject; because it is an ‘unknown’ parameter, it must also be estimated from the data at hand. Although the focus of this paper has been on the proportional odds logistic regression model, we note that, if our approach is applied to the multinomial logistic regression model for nominal (unordered) responses as outlined in Appendix B, no additional nuisance parameters need to be included in the model in the Poisson likelihood; the p_{ij} s in equation (1) are simply the multinomial model probabilities. However, the resulting expression for the pseudoresponse for the multinomial logistic model is not as simple as in the case of the proportional odds model.

Finally, the results of the simulations demonstrate that the method proposed can greatly reduce the finite sample bias of ML for estimating the regression parameters of the proportional odds logistic regression model. Interestingly, even in simulations where none of the response categories were rare, the standard ML approach was found to have substantial bias in small samples. However, because of the broad range of possible data configurations, it is difficult to draw definitive conclusions from the results of the simulation studies. Nonetheless, in the simulations that are reported here, the bias-corrected method performs discernibly better than the standard likelihood approach, suggesting that the bias-corrected method could be adopted as a first-line choice in regression analyses of ordinal outcomes.

Acknowledgements

We are grateful for the support provided by grants MH 054693 and CA 160679 from the US National Institutes of Health. We thank the Joint Editor, Associate Editor and referees for their helpful comments and suggestions on a previous version of the manuscript.

Appendix A: Multinomial likelihood equations

Here we show that the score equations for β from the multinomial likelihood are identical to the score equations given by equations (2) in Section 2. We also show that the expected information is identical under the multinomial and Poisson likelihood formulations (the observed information will be the same since the score equations are the same).

We denote the $J \times 1$ vector of multinomial indicator random variables for subject i as $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})'$. Although the Y_{ij} s sum to 1 for each i , we adopt the convention of McCullagh and Nelder (1989) and include all J indicators in the outcome vector. Further, $E(\mathbf{Y}_i | \mathbf{x}_i) = \mathbf{p}_i = (p_{i1}, \dots, p_{iJ})'$, and the variance–covariance matrix of \mathbf{Y}_i equals

$$V_i = \text{var}(\mathbf{Y}_i) = \text{diag}(\mathbf{p}_i) - \mathbf{p}_i \mathbf{p}_i'$$

where $\text{diag}(\mathbf{p}_i)$ is a diagonal matrix with the elements of \mathbf{p}_i on the diagonal. Because $\sum_{j=1}^J Y_{ij} = 1$, $\text{var}(\mathbf{Y}_i)$ has rank $J - 1$. McCullagh and Nelder (1989), page 167, defined the generalized inverse of V_i as

$$V_i^- = \text{diag}(\mathbf{p}_i)^{-1},$$

i.e. a diagonal matrix with $1/p_{ij}$ on the diagonal. This generalized inverse has rank J and satisfies the property that

$$V_i V_i^- V_i = V_i.$$

Then, under any model with the constraint $\sum_{j=1}^J p_{ij} = 1$, the multinomial ML equations for β are

$$u(\hat{\beta}) = \sum_{i=1}^n \hat{D}_i \hat{V}_i^{-1} (\mathbf{Y}_i - \hat{\mathbf{p}}_i) = \sum_{i=1}^n \sum_{j=1}^J \hat{D}_{ij} \hat{p}_{ij}^{-1} (y_{ij} - \hat{p}_{ij}) = 0, \tag{8}$$

where $D_{ij} = \partial p_{ij}(\beta) / \partial \beta$, and the j th column of D_i equals D_{ij} (see, for example, McCullagh and Nelder (1989), pages 171–172). Note that these are identical to the Poisson score equations given by equations (2) in Section 2.

The observed information matrix can be written as

$$\sum_{i=1}^n (D_i V_i^{-1} D_i' - A_i)$$

where the k th column of A_i has typical element

$$A_{ik} = \left(\frac{\partial(D_i V_i^{-1})}{\partial \beta_k} \right) (\mathbf{Y}_i - \mathbf{p}_i). \tag{9}$$

Under both the multinomial and Poisson formulation outlined in Section 2, the first moment of \mathbf{Y}_i equals \mathbf{p}_i , i.e. $E(\mathbf{Y}_i - \mathbf{p}_i) = 0$. Therefore,

$$E(A_{ik}) = \left(\frac{\partial(D_i V_i^{-1})}{\partial \beta_k} \right) E(\mathbf{Y}_i - \mathbf{p}_i) = 0,$$

and $E(A_i) = 0$. Thus, the expected (Fisher) information matrix equals

$$\sum_{i=1}^n \{D_i V_i^{-1} D_i' - E(A_i)\} = \sum_{i=1}^n (D_i V_i^{-1} D_i')$$

under both the multinomial and the Poisson formulation outlined in Section 2. This establishes that bias correction, which is a function of $\text{tr}\{\text{var}(\hat{\beta}) D_{ij}^2\}$ where $\text{var}(\hat{\beta})$ can be based on either the observed or the expected information, is the same whether based on the multinomial likelihood or by substituting a Poisson likelihood subject to the model constraining $\sum_{j=1}^J p_{ij} = 1$.

Appendix B: Multinomial logistic regression

Here we briefly discuss implementation of our first-order bias correction approach for a multinomial logistic regression model for nominal (unordered) responses. We also show that our approach is not appropriate when the multinomial logistic regression model is expressed in terms of a Poisson log-linear model with subject-specific effects.

In a slight departure from the notation in previous sections, the multinomial logistic regression model can be written as

$$p_{ij} = \text{pr}(Y_i = j | \mathbf{x}_i, \beta) = \frac{\exp(\mathbf{x}'_{ij} \beta_j)}{\sum_{j=1}^J \exp(\mathbf{x}'_{ij} \beta_j)} \quad j = 1, \dots, J. \tag{10}$$

where \mathbf{x}_{ij} is the covariate vector corresponding to multinomial level j (ordinarily, \mathbf{x}_{ij} contains the covariates \mathbf{x}_i plus an indicator for the intercept for level j) and β_j are the regression parameters corresponding to level j (often β_j is set to 0 for identifiability). Since this multinomial logistic model has the constraint

$$\sum_{j=1}^J p_{ij} = \sum_{j=1}^J \exp(\mathbf{x}'_{ij} \beta_j) / \sum_{j=1}^J \exp(\mathbf{x}'_{ij} \beta_j) = 1$$

satisfied by definition, we can use the Poisson likelihood approach that is discussed in this paper with p_{ij} specified as in equation (10). The pseudoresponses have the same form as in equation (3), with $a_{ij} = 0.5 \text{tr}\{\text{var}(\hat{\beta}) D_{ij}^2\}$. Although it has a closed form and can be calculated in a matrix software package (e.g. R or SAS PROC IML), the resulting expression for the pseudoresponse for the multinomial logistic model is not as simple as in the case of the proportional odds model that is the focus of this paper.

Next, suppose that the multinomial logistic regression model is written as a Poisson log-linear model. In particular, the Poisson log-linear model for p_{ij} is

$$\log(p_{ij}) = \beta_{0i} + \mathbf{x}'_{ij} \beta_j,$$

where β_{0i} is an effect for the i th subject. Using properties of sufficient statistics for a log-linear model with a Poisson-distributed outcome, these subject-specific β_{0i} s constrain $\sum_{j=1}^J p_{ij} = \sum_{j=1}^J Y_{ij}$. In this paper (before applying the bias correction approach), we assume the general situation where all subjects have unique covariates so that $\sum_{j=1}^J Y_{ij} = 1$. Then, for the Poisson log-linear model, since $\sum_{j=1}^J Y_{ij} = 1$, it follows that $\sum_{j=1}^J p_{ij} = 1$. In this case, it is easily shown that the estimate of β_j will be the same from directly maximizing the multinomial likelihood or by fitting the Poisson log-linear model.

However, our implementation of the bias correction cannot be applied to the Poisson log-linear model. The reason is as follows. If we directly attempted to apply our bias correction approach to a Poisson log-linear model, in the iterative bias correction algorithm, we have the pseudoresponse

$$Y_{ij}^* = Y_{ij} + a_{ij}$$

where $a_{ij} > 0$, so

$$\sum_{j=1}^J p_{ij} = \sum_{j=1}^J Y_{ij}^* > 1.$$

Thus, our implementation of the bias correction, which requires $\sum_{j=1}^J p_{ij} = 1$, cannot be applied to a Poisson log-linear model version of the multinomial logistic regression model. For the Poisson log-linear formulation of the multinomial logistic regression model, Kosmidis and Firth (2011) have given an elegant approach to implementing the first-order bias correction.

References

- Agresti, A. A. (2010) *Analysis of Ordinal Categorical Data*, 2nd edn. Hoboken: Wiley.
- Birch, M. W. (1963) Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc. B*, **25**, 220–233.
- Box, M. J. (1971) Bias in nonlinear estimation (with discussion). *J. R. Statist. Soc. B*, **33**, 171–201.
- Bull, S. B., Lewinger, J. B. and Lee, S. S. F. (2007) Confidence intervals for multinomial logistic regression in sparse data. *Statist. Med.*, **26**, 903–918.
- Bull, S. B., Mak, C. and Greenwood, C. (2002) A modified score function estimator for multinomial logistic regression in small samples. *Computnl Statist. Data Anal.*, **39**, 57–74.
- Clogg, C. C., Rubin, D. B., Schenker, N., Schultz, B. and Weidman, L. (1991) Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *J. Am. Statist. Ass.*, **86**, 68–78.
- Falcoz, P. E., Conti, M., Brouchet, L., Chocron, S., Puyraveau, M., Mercier, M., Etievent, J. P. and Dahan, M. (2007) The Thoracic Surgery Scoring System (Thoracoscore): risk model for in-hospital death in 15,183 patients requiring thoracic surgery. *J. Thorac. Cardvasc. Surg.*, **133**, 325–332.
- Firth, D. (1993) Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.
- Gawande, A. A., Kwaan, M. R., Regenbogen, S. E., Lipsitz, S. R. and Zinner, M. J. (2007) An Apgar score for surgery. *J. Am. Coll. Surg.*, **204**, 201–208.
- Guru, V., Frenes, S. E., Austin, P. C., Blackstone, E. H. and Tu, J. V. (2006) Gender differences in outcomes after hospital discharge from coronary artery bypass grafting. *Circulation*, **113**, 507–516.
- Hauck, W. W. and Donner, A. (1977) Wald's test as applied to hypotheses in logit analysis. *J. Am. Statist. Ass.*, **72**, 851–853.
- Heinze, G. and Schemper, M. (2002) A solution to the problem of separation in logistic regression. *Statist. Med.*, **21**, 2409–2419.
- Kosmidis, I. and Firth, D. (2009) Bias reduction in exponential family nonlinear models. *Biometrika*, **96**, 793–804.
- Kosmidis, I. and Firth, D. (2011) Multinomial logit bias reduction via the Poisson log-linear model. *Biometrika*, **98**, 755–759.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
- Nguyen, L. L., Hevelone, N., Rogers, S. O., Bandyk, D. F., Clowes, A. W., Moneta, G. L., Lipsitz, S. and Conte, M. S. (2009) Disparity in outcomes of surgical revascularization for limb salvage: race and gender are synergistic determinants of vein graft failure and limb loss. *Circulation*, **119**, 123–130.
- Owens, W. D., Felts, J. A. and Spitznagel, Jr, E. L. (1978) ASA physical status classifications: a study of consistency of ratings. *Anesthesiology*, **49**, 239–243.
- SAS Institute (2010) SAS Online Doc, version 9.2. SAS Institute, Cary. (Available from <http://www.sas.com>.)
- Williams, O. D. and Grizzle, J. E. (1972) Analysis of contingency tables having ordered response categories. *J. Am. Statist. Ass.*, **67**, 55–63.