

The Log Multinomial Regression Model for Nominal Outcomes with More than Two Attributes

L. Blizzard^{*,1} and D. W. Hosmer²

¹ Menzies Research Institute, University of Tasmania, Private Bag 23, Hobart, TAS 7001, Australia

² Department of Public Health, University of Massachusetts, Amherst, MA 01003-9304, USA

Received 20 November 2006, revised 5 June 2007, accepted 29 June 2007

Summary

An estimate of the risk or prevalence ratio, adjusted for confounders, can be obtained from a log binomial model (binomial errors, log link) fitted to binary outcome data. We propose a modification of the log binomial model to obtain relative risk estimates for nominal outcomes with more than two attributes (the “log multinomial model”). Extensive data simulations were undertaken to compare the performance of the log multinomial model with that of an expanded data multinomial logistic regression method based on the approach proposed by Schouten et al. (1993) for binary data, and with that of separate fits of a Poisson regression model based on the approach proposed by Zou (2004) and Carter, Lipsitz and Tilley (2005) for binary data. Log multinomial regression resulted in “inadmissible” solutions (out-of-bounds probabilities) exceeding 50% in some data settings. Coefficient estimates by the alternative methods produced out-of-bounds probabilities for the log multinomial model in up to 27% of samples to which a log multinomial model had been successfully fitted. The log multinomial coefficient estimates generally had lesser relative bias and mean squared error than the alternative methods. The practical utility of the log multinomial regression model was demonstrated with a real data example. The log multinomial model offers a practical solution to the problem of obtaining adjusted estimates of the risk ratio in the multinomial setting, but must be used with some care and attention to detail.

Key words: Log binomial regression; Log link; Multinomial likelihood; Risk ratio.

1 Introduction

Fitting a generalized linear model (McCullough and Nelder, 1989) with a non-canonical logarithmic link to binomial data makes it possible to directly estimate risk ratios in prospective studies and prevalence ratios in cross-sectional studies. This is the log binomial model (Wacholder, 1986; Blizzard and Hosmer, 2006). Its use is advocated (Zochetti, Consinni and Bertazzi, 1995; Skov et al., 1998; McNutt et al., 2003; Deddens, Petersen and Lei, 2003; Spiegelman and Hertzmark, 2005; Barros and Hirakata, 2006; Lumley, Kronmal and Ma, 2006) particularly when the binary outcome is not rare. The odds ratio produced by logistic regression approximates the risk or prevalence ratio, but the approximation is adequate only when the outcome is rare in all exposure and confounder categories (Greenland, 1987) and is liable to misinterpretation and misreporting when it is not (Katz, 2006).

For nominal outcome data with more than two levels, the choice of modelling methods is limited presently to a modification of the binary logistic regression model. An early development in the business and econometric literature in this respect was the discrete choice model proposed by McFadden (1974). It is termed the multinomial, polychotomous or polytomous logistic regression model in the

* Corresponding author: e-mail: Leigh.Blizzard@utas.edu.au, Phone: +61 3 6226 7719, Fax: +61 3 6226 7704

health and life sciences. For example, consider a follow-up study of children into adulthood with daily tobacco smoking as the outcome. The outcome categories are never-smoker, former smoker and current smoker. Possible covariates include sex, childhood attitudes to smoking, and continuous factors such as age. Multinomial logistic regression makes it possible to approximate the relative risk of being a former smoker and the relative risk of being a current smoker at follow-up with adjustment for continuous covariates.

In principle, it should be possible to modify the log binomial model to obtain relative risk estimates for data of this type. We are not aware of this having been attempted previously. In this paper, we examine some of the technical issues that arise when one attempts to fit a model with a log link to multinomial data. We refer to it as the log multinomial model. The paper is organised as follows. In Section 2, the log multinomial regression model is introduced and likelihood-based estimation methods are reviewed. Two alternative methods, one based on the approach of Schouten et al. (1993) for binary data, and another involving separate fits of a Poisson regression model, are proposed in Section 3. In Section 4, we present the results of a simulation study of the performance of the estimation methods. An application to real data is provided in Section 5. Finally, the implications are summarised in Section 6.

2 The Log Multinomial Regression Model

Suppose that individuals in an infinitely large population possess only one of the $(J + 1)$ attributes $A_0, A_1, A_2, \dots, A_J$. Let random variable Y denote the outcome of sampling from this population, where $Y = j$ if attribute A_j is observed, and let $Y_j, j = 0, 1, 2, \dots, J$ denote binary (0/1) random variables indicating which attribute is observed, with $\sum_{j=0}^J Y_j = 1$. Assume the mean π_j of each Y_j depends upon a linear combination of the observed values x_1, x_2, \dots, x_K of K non-constant covariates X_1, X_2, \dots, X_K . Denote the linear predictor as $\mathbf{x}'\boldsymbol{\beta}_j = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jK}x_K$ where $\mathbf{x}' = (1, x_1, x_2, \dots, x_K)$, and where $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \beta_{j2}, \dots, \beta_{jK}), j = 0, 1, 2, \dots, J$ are parameters to be estimated. Assume further that the $\pi_j = \pi_j(\mathbf{x})$ follow the simple exponential form $\pi_j(\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta}_j)$. The requirement $\sum_{j=0}^J \pi_j = 1$ renders as redundant the modelling of one outcome category (say the first, corresponding to $j = 0$). Then the log multinomial regression model is:

$$\Pr(Y_j = 1 | \mathbf{x}) = \pi_j(\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta}_j), \quad j = 1, 2, \dots, J \quad (1)$$

with $Y_0 = 1 - \sum_{q=1}^J Y_q$ and $\Pr(Y_0 = 1 | \mathbf{x}) = 1 - \sum_{q=1}^J \Pr(Y_q = 1 | \mathbf{x})$, and with the random and systematic components linked by the logarithmic transform $\ln[\pi_j(\mathbf{x})] = \mathbf{x}'\boldsymbol{\beta}_j$. Finally we assume there are n independent observations of the multinomial outcome variable Y and the covariates, and denote the observed data as $(\mathbf{y}_i, \mathbf{x}_i), i = 1, 2, \dots, n$ where $\mathbf{y}_i' = (y_{i0}, y_{i1}, y_{i2}, \dots, y_{iJ})$.

The conditional log likelihood function is:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \left\{ \left(1 - \sum_{q=1}^J y_{iq}\right) \ln \left[1 - \sum_{q=1}^J \pi_q(\mathbf{x}_i)\right] + \sum_{q=1}^J y_{iq} \ln [\pi_q(\mathbf{x}_i)] \right\} \quad (2)$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0 \boldsymbol{\beta}_1 \boldsymbol{\beta}_2 \dots \boldsymbol{\beta}_J)$. The likelihood equations for the parameters are found by taking the first partial derivatives of $L(\boldsymbol{\beta})$ with respect to the $J(K + 1)$ parameters $\beta_{jk}, j = 1, 2, \dots, J, k = 0, 1, 2, \dots, K$. These $J(K + 1) \times 1$ score equations may be written as the sum of n gradient vectors $\mathbf{g}_i(\boldsymbol{\beta})$ of dimension $J(K + 1) \times 1$ containing the scores for i -th subject:

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n [\mathbf{g}_i(\boldsymbol{\beta})].$$

Under standard maximum likelihood theory, the observed information matrix $\mathbf{I}(\hat{\boldsymbol{\beta}})$ is the negative of the $J(K + 1) \times J(K + 1)$ matrix of second partial derivatives evaluated at the likelihood-maximising solution $\hat{\boldsymbol{\beta}}$, and an estimator of the covariance matrix of the maximum likelihood estimator is its inverse:

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = [\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}})]^{-1}. \quad (3)$$

The elements of $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$ depend on the observed value of the outcome variable, its fitted values and the covariate values. In contrast, the elements of the observed information matrix for the multinomial logistic model are independent of the observed values of the outcome variable. This difference is a consequence of using the log link in place of the logit link. An “information sandwich” estimator is used to provide robustness to model misspecification. The “meat” of the sandwich is the $J(K+1) \times J(K+1)$ matrix $\hat{\mathbf{M}}$ that is based on the outer product of the gradient vector evaluated at $\hat{\boldsymbol{\beta}}$:

$$\hat{\mathbf{M}} = \frac{n}{n-1} \sum_{i=1}^n [\mathbf{g}_i(\hat{\boldsymbol{\beta}})] [\mathbf{g}_i(\hat{\boldsymbol{\beta}})]'.$$

The robust version using the observed information matrix is:

$$\hat{\mathbf{R}} = [\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}})]^{-1} \times \hat{\mathbf{M}} \times [\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}})]^{-1}. \quad (4)$$

An issue in fitting the log multinomial regression model is that the permissible parameter space for a maximum likelihood estimator of $\boldsymbol{\beta}$ is constrained to the subspace of $R^{J(K+1)}$ that generates fitted probabilities that lie between 0 and 1. Two of the adverse outcomes in attempting to fit the model in practice are that the iterative methods used by the software package may fail to converge, and that a converged solution may not be admissible. Here we define an admissible solution as a value of $\boldsymbol{\beta}$ such that $\mathbf{x}'_i \boldsymbol{\beta}_{ij} < 0$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, J$ and hence $0 < \hat{\pi}_j(\mathbf{x}_i) < 1$, $i = 1, 2, \dots, n$, $j = 0, 1, 2, \dots, J$. This requires that $\sum_{j=1}^J \hat{\pi}_j(\mathbf{x}_i) < 1$, $i = 1, 2, \dots, n$. We discuss this more extensively in later sections.

3 Two Alternative Models for Modelling Multinomial Outcomes

3.1 The expanded data logistic regression approach of Schouten et al.

To estimate relative risk for binomial data using logistic regression, Schouten et al. (1993) proposed expanding the data so that comparison group includes all subjects in the sample rather than the “non-cases” only. To do this for binary outcomes $Y = 0$ or 1, a new binary indicator variable D is defined with value $d = y$, and the data for subjects with $y = 1$ are replicated but with the additional observations in the augmented data assigned the value $d = 0$. This logistic regression model applied to the expanded data yields consistent estimators of the log binomial slope coefficients. They also derive a consistent “information sandwich” type of estimator of the covariance matrix of the estimated coefficients.

Modifying this approach for multinomial data is equivalent to maximising over the original sample the quasi-likelihood:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \left\{ \left(1 + \sum_{\varphi=1}^J y_{i\varphi} \right) \left[\left(1 - \sum_{\varphi=1}^J y_{i\varphi} \right) \ln \left(1 - \sum_{\varphi=1}^J \theta_{\varphi}(\mathbf{x}_i) \right) + \sum_{\varphi=1}^J y_{i\varphi} \ln [\varphi_{\varphi}(\mathbf{x}_i)] \right] \right\}$$

with $\mathbf{x}'_i \boldsymbol{\alpha}_j = \alpha_{j0} + \alpha_{j1}x_{i1} + \beta_{j2}x_{i2} + \dots + \alpha_{jK}x_{iK}$, $\mathbf{x}'_i \boldsymbol{\alpha}_0 = 0$ and $\exp(\mathbf{x}'_i \boldsymbol{\alpha}_0) = 1$, and where the $\theta_j(\mathbf{x})$ follow the multinomial logistic regression model:

$$\Pr(Y_{ij} = 1 \mid \mathbf{x}_i) = \theta_j(\mathbf{x}_i) = \frac{\exp(\mathbf{x}'_i \boldsymbol{\alpha}_j)}{1 + \sum_{\varphi=1}^J \exp(\mathbf{x}'_i \boldsymbol{\alpha}_j)}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, J.$$

An estimator of the covariance matrix of the estimated coefficients $\hat{\boldsymbol{\alpha}}_j$ can be obtained using an “information sandwich” analogous to that described by Schouten et al. (1993) for the binary case. The estimator of the variance-covariance matrix of the expanded data multinomial logistic model is:

$$\hat{\mathbf{\Lambda}} = \hat{\mathbf{V}} \times \hat{\mathbf{H}} \times \hat{\mathbf{V}} \quad (5)$$

where $\hat{\mathbf{H}}$ denotes the outer product of the gradient vectors. In this case, $\hat{\mathbf{H}}$ has the form:

$$\hat{\mathbf{H}} = \sum_{i=1}^n [\mathbf{r}_i(\hat{\boldsymbol{\alpha}})] [\mathbf{r}_i(\hat{\boldsymbol{\alpha}})]'$$

where $\mathbf{r}_i(\hat{\boldsymbol{\alpha}})$ is a gradient vector of dimension $J(K+1)$ containing the estimated scores for i -th subject computed over the original sample with typical element:

$$\hat{r}_{ijk} = \left[y_{ij} - \left(\sum_{q=1}^J y_{iq} + 1 \right) \hat{\theta}_j(\mathbf{x}_i) \right] x_{ik}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, J, \quad k = 0, 1, 2, \dots, K.$$

Estimates $\hat{\theta}_j(\mathbf{x}_i)$ of the parameters of the log multinomial model, and of their estimated standard errors, can be obtained in this way from nearly any set of data. The solution is not guaranteed to be admissible for the log multinomial model, however. This is an issue we investigate in Section 4.

3.2 Separate Poisson regression models

To estimate relative risk for binomial data, McNutt et al. (2003) investigated the use of the Poisson regression model with log link and Poisson errors that has the same form as the log binomial model:

$$\Pr(Y = 1 | \mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\gamma}).$$

The coefficient estimates $\hat{\boldsymbol{\gamma}}$ consistently estimate the coefficients from the log binomial model (Carter et al., 2005). This approach requires no data modification, and can be easily performed using widely available software such as SAS and Stata. It produces confidence intervals that are too wide because the Poisson errors are over-estimates of the binomial errors when the outcome is not rare. Zou (2004) proposed correcting for this by using the information sandwich estimator to obtain variance estimates that are robust to the error misspecification. Carter et al. (2005) showed that by using the information sandwich estimator of the covariance matrix of the estimated coefficients following the Poisson regression fit, one obtains a consistent estimator of the covariance matrix of estimated coefficients from a log binomial fit.

Whilst there is no Poisson model for multinomial data, it is possible to fit separate binary Poisson regression models to the data for J of the $J+1$ outcomes. If outcome $j = 0$ is chosen as the redundant category, this involves fitting the separate Poisson regression models:

$$\Pr(Y_{ij} = 1 | \mathbf{x}) = \exp(\mathbf{x}'_i \boldsymbol{\gamma}_j), \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, J.$$

Whilst cumbersome, this procedure is straightforward and we used it to provide a further benchmark against which to compare the estimates from the log multinomial model. Because the mean of the Poisson regression model is not constrained to be less than one, however, the resulting estimated coefficients may yield inadmissible fitted values for the log binomial model probabilities. This is investigated in the simulations of Section 4.

4 Simulation Study

In this section, we describe simulation studies conducted to assess practical aspects of fitting the log multinomial model and to evaluate its performance. Results are presented using three summary statistics. The first two are average percent relative bias, $(100/L) \sum_{l=1}^L [(\hat{\beta}_{jk}^{(l)} - \beta_{jk})/\beta_{jk}]$, $j = 1, 2, \dots, J$; $k = 0, 1, \dots, K$, and average percent mean squared error, $(100/L) \sum_{l=1}^L [(\hat{\beta}_{j1}^{(l)} - \beta_{j1})^2 + \widehat{\text{Var}}(\hat{\beta}_{j1}^{(l)})]$, $j = 1, 2$; $k = 0, 1, \dots, K$, where $\hat{\beta}_{jk}^{(l)}$ is the coefficient estimated from the l -th of L replications of the data and β_{jk} is its design ("true") value. The third is the observed coverage of confidence intervals calculated as $\hat{\beta}_{jk}^{(l)} \pm z_{1-\alpha/2} \times \widehat{\text{SE}}(\hat{\beta}_{jk}^{(l)})$, $j = 1, 2, \dots, J$; $l = 1, 2, \dots, L$ where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard normal distribution, and where $\alpha = 0.05$ for 95% confidence intervals. The estimated standard errors $\widehat{\text{SE}}(\hat{\beta}_{jk}^{(l)})$ were produced by the robust version (4) of the observed information matrix for the log multinomial model, the covariance matrix (5) robust to duplication of observations for the expanded data multinomial logistic model, and the information sandwich estimator for the separate Poisson regression fits.

All calculations were performed using Stata (Stata Corp., 2003) with modified Newton-Raphson optimisation performed using the *ml* command (Gould, Pitbaldo and Sribney, 2003). In standard implementation of its estimation algorithm, evaluation of the terms $y_{iq} \times \ln [\pi_q(\mathbf{x}_i)]$ in the log likelihood:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \left\{ \sum_{q=0}^J y_{iq} \times \ln [\pi_q(\mathbf{x}_i)] \right\}$$

requires evaluation of $\ln [\pi_q(\mathbf{x}_i)]$ even if $y_{iq} = 0$. This step involves needless attempted evaluation of $\ln [\pi_0(\mathbf{x}_i)]$ when $\hat{\pi}_0(\mathbf{x}_i) = 1 - \sum_{q=1}^J \hat{\pi}_q(\mathbf{x}_i) \leq 0$. To avoid this, the algorithm was written in a manner that ensured each $y_{iq} \times \ln [\pi_q(\mathbf{x}_i)]$ was evaluated only when $y_{iq} = 1$. This reduced non-convergence of the estimation algorithm to a very rare occurrence, and replaced non-convergent solutions with convergent but inadmissible solutions with $\hat{\pi}_0(\mathbf{x}_i) = 1 - \sum_{q=1}^J \hat{\pi}_q(\mathbf{x}_i) \leq 0$. This simple alteration enabled us to study and report results for inadmissible solutions.

4.1 Simulations I

The basic design was a univariable model for a three-category ($J = 3$) multinomial outcome variable Y and a single ($K = 1$) covariate X distributed uniformly on the interval $[-6, a]$. The settings are summarised in Table 1.

Binary (0/1) indicator variables Y_j , $j = 0, 1, 2$ were used to denote the outcomes with Y_0 the redundant category. The intercept (β_{20}) and slope (β_{21}) coefficients in the model for Y_2 were chosen such that $\Pr(Y_2 = 1 \mid x = -6) = 0.01$ and $\Pr(Y_2 = 1 \mid x = 0) = b$ where $b = 0.1, 0.3, 0.5$ and 0.7 . The upper bound a of the uniform distribution was then chosen so that the marginal probability of response:

$$\Pr(Y_2 = 1) = \int_{-6}^a e^{\beta_{20} + \beta_{21}x} \frac{1}{a - (-6)} dx = \frac{e^{\beta_{20} + \beta_{21}a} - e^{\beta_{20} + \beta_{21}(-6)}}{(a + 6) \times \beta_{21}}$$

Table 1 Design of the first set of simulations.

Setting	Outcome category $j = 1, 2$	β_{j0}	β_{j1}	a^*	$\Pr(Y_j = 1)$ at $x = -6$	$\Pr(Y_j = 1)$ at $x = 0$	$\Pr(Y_j = 1)$ at $x = a$	Marginal probability $\Pr(Y_j = 1)$
1	1	-2.6323	0.3288	4.0	0.01	0.0719	0.2679	0.0784
	2	-2.3026	0.3838	4.0	0.01	0.1000	0.4642	0.1183
2	1	-2.6271	-0.3784	2.0	0.70	0.0723	0.0339	0.2200
	2	-1.2040	0.5669	2.0	0.01	0.3000	0.9322	0.2033
3	1	-1.8970	0.4514	1.0	0.01	0.1500	0.2356	0.0714
	2	-1.2040	0.5669	1.0	0.01	0.3000	0.5288	0.1307
4	1	-3.3975	-0.5068	1.0	0.70	0.0335	0.0202	0.1916
	2	-0.6931	0.6520	1.0	0.01	0.5000	0.9597	0.2081
5	1	-1.3863	0.5365	0.0	0.01	0.2500	0.2500	0.0746
	2	-0.6931	0.6520	0.0	0.01	0.5000	0.5000	0.1253
6	1	-6.1510	-0.9657	0.5	0.70	0.0021	0.0013	0.1113
	2	-0.3567	0.7081	0.5	0.01	0.7000	0.9974	0.2145
7	1	-1.0748	0.5884	-0.5	0.01	0.3414	0.2544	0.0755
	2	-0.3567	0.7081	-0.5	0.01	0.7000	0.4913	0.1236

* Maximum value of the covariate $X \sim U(-6, a)$.

Table 2 Summary statistics describing the frequency of problems encountered when fitting models to the log multinomial data for settings 1–7.

Setting	Admissible samples †			Samples per 1000 admissible LnMM fits ‡	Inadmissible §	
	LnMM*	EMLM*	Poisson*		EMLM*	Poisson*
1	99.9	99.6	99.9	1001	0.4	0.1
2	67.8	63.9	66.3	1476	24.0	19.2
3	99.9	99.2	99.7	1001	0.8	0.3
4	59.9	59.5	61.4	1669	25.3	20.5
5	99.9	99.4	99.8	1001	0.6	0.2
6	45.4	51.6	52.2	2204	27.7	23.5
7	99.9	99.6	99.9	1001	0.3	0.0

* LnMM = log multinomial regression model, EMLM = expanded multinomial logistic regression model, Poisson = separate Poisson regression models.

† Percentage of samples for which the estimation algorithm converged and produced admissible probabilities for the log multinomial model.

‡ Number of samples required to produce each 1000 admissible fits of a LnMM.

§ Percentage of the samples producing admissible fits of a LnMM for which the estimated coefficients from the EMLM or the separate Poisson regression models produced at least one inadmissible log multinomial model probability.

was 0.1 or 0.2 (approximately, after the values of a were rounded). The intercept β_{10} and slope coefficient β_{11} from the model for Y_1 were then chosen such that $\Pr(Y_1 = 1 | x = -6) = c$ where $c = 0.01$ or 0.70 , and $\Pr(Y_1 = 1 | x = a) = 0.5 \times [1 - \Pr(Y_2 = 1 | x = a)]$. Finally, outcomes were assigned by drawing a uniform random deviate (v) from a $U(0, 1)$ distribution, and setting $y = 0$ if $v \leq \Pr(Y = 0 | x)$, $y = 1$ if $\Pr(Y = 0 | x) < v \leq \Pr(Y = 0 | x) + \Pr(Y = 1 | x)$, or $y = 2$ otherwise. We drew samples of size 500 because datasets of size $n = 500$ are commonly encountered in practice, with an adequate number of replications to produce at least 1000 inadmissible solutions in settings 2, 4 and 6.

The results of fitting the three models are shown in Table 2. We expected convergence and/or admissibility problems in settings 2, 4 and 6 where the greater of the maximum design probabilities $\Pr(Y_2 = 1 | x = a)$ was 0.93 or greater, and the sum of the maximum design probabilities was at least 0.95, but not in the remaining settings where the sum of the maximum design probabilities was lower (around 0.75). As expected, fitting problems occurred in settings 2, 4 and 6 and their frequency increased with the size and sum of the maximum design probabilities. The samples that produced fitting problems for each model did not correspond exactly, with the expanded data multinomial logistic regression model and the separate Poisson regression models producing inadmissible probabilities for the log multinomial model in 19–28 percent of samples successfully fitted by a log multinomial regression model.

The results of estimating the coefficient of the uniformly distributed, continuous covariate are shown in Table 3. Bias and mean squared error were small in the admissible samples and, whilst the differences between models were minor, the log multinomial estimator generally had least bias and mean squared error. The expanded data multinomial logistic estimator had the poorest performance in these respects. The confidence interval coverage for each model straddled 95 percent, with those produced by the log multinomial estimator most often closest to the target. Given the design settings, the inadmissible samples for settings 2, 4 and 6 were produced by inflated estimates of the slope coefficient of the continuous covariate in the second outcome category (β_{21}). In consequence, the bias and mean squared error of the estimates of that parameter were substantially higher in the inadmissi-

Table 3 Results of estimating the coefficients of the uniform covariate in settings 1–7.

	Outcome category	LnMM*			EMLM*			Poisson		
Setting	$j = 1, 2$	Bias †	MSE ‡	Cover §	Bias †	MSE ‡	Cover §	Bias †	MSE ‡	Cover §
Admissible samples ¶										
1	1	2.04	0.94	94.6	2.56	0.99	94.3	2.39	0.97	94.4
	2	0.98	0.63	94.7	1.46	0.69	94.5	1.31	0.66	94.6
2	1	0.04	0.36	94.8	−0.33	0.42	94.7	−0.02	0.39	94.6
	2	−2.09	0.45	95.9	−3.64	0.61	95.7	−3.64	0.54	95.7
3	1	1.96	2.08	94.2	2.37	2.16	93.9	2.24	2.12	94.0
	2	0.66	1.18	94.5	1.08	1.31	94.7	0.96	1.25	94.5
4	1	0.90	0.65	94.6	0.25	0.76	94.3	0.53	0.70	94.3
	2	−2.12	0.58	95.5	−3.65	0.77	96.2	−3.02	0.69	96.1
5	1	1.48	2.77	94.2	1.86	2.88	94.1	1.78	2.83	94.1
	2	1.05	1.66	95.1	1.50	1.83	95.2	1.35	1.74	95.3
6	1	1.24	2.87	95.4	1.33	3.33	94.6	1.34	3.09	94.6
	2	−2.96	0.66	95.7	−4.65	0.90	94.9	−4.13	0.80	94.8
7	1	1.79	3.19	94.8	2.31	3.34	94.8	2.13	3.27	94.8
	2	0.93	2.02	95.0	1.48	2.25	94.8	1.29	2.13	94.9
Inadmissible samples ¶										
2	1	1.06	0.35	95.7	2.34	0.41	95.4	1.44	0.38	95.5
	2	9.72	0.89	89.6	9.42	0.90	91.1	8.50	0.77	91.4
4	1	0.83	0.65	93.9	2.43	0.77	93.9	1.60	0.71	94.0
	2	8.40	1.04	90.2	8.31	1.08	92.4	7.31	0.92	92.5
6	1	1.46	2.99	94.7	2.36	3.38	93.5	1.80	3.10	94.6
	2	7.44	1.15	91.3	7.10	1.13	93.5	6.28	0.97	94.3

* LnMM = log multinomial regression model, EMLM = expanded multinomial logistic regression model, Poisson = separate Poisson regression models.

† Average percent relative bias of the slope estimator.

‡ Average percent mean squared error of the slope estimator.

§ Percentage of 95 percent confidence intervals covering the true value of the coefficient.

¶ Admissible samples are converged solutions those that produced admissible probabilities for the log multinomial model. Inadmissible samples are converged solutions that did not.

ble samples than in the admissible samples, and bias reached nearly 10 percent for the log multinomial estimator in setting 2. With only modest differences in the standard error estimates, the increased bias was the main contributor to poorer confidence interval coverage. There was little to choose between the three models. Results are not provided for settings 1, 3, 5 and 7 because inadmissible samples were rare in those settings.

4.2 Simulations II

A second set of simulations was performed with models for a three-category ($J = 3$) multinomial outcome Y with two covariates: a dichotomous covariate $D \sim B(p)$ having Bernoulli probability $p = 0.20, 0.35$ and 0.50 in settings 8–10 respectively, and a uniformly distributed continuous covariate U associated with D via density function $f(U | D = d) = U(-6 + 2 \times d, 2 + 2 \times d)$. In the model for Y_2 , we set the intercept coefficient as $\beta_{20} = \ln(0.3)$ and the slope coefficient for the dichotomous

Table 4 Design of the second set of simulations.

Setting	Outcome category $j = 1, 2$	Intercept β_{j0}	Binary covariate β_{j1}	Continuous covariate β_{j2}	Bernoulli probability p	$\Pr(Y_j = 1)$ at $d = 1, u = 4$	Marginal probability $\Pr(Y_j = 1)$
8	1	-1.7635	-0.5192	-0.1784	0.20	0.05	0.2352
	2	-1.2040	0.4055	0.1733	0.20	0.90	0.2801
9	1	-1.6519	-0.8534	-0.1226	0.35	0.05	0.1954
	2	-1.2040	0.5596	0.1348	0.35	0.90	0.3490
10	1	-1.5871	-1.1534	-0.0639	0.50	0.05	0.1501
	2	-1.2040	0.6931	0.1014	0.50	0.90	0.4341

covariate as $\beta_{21} = \ln(1.50)$, $\ln(1.75)$ and $\ln(2.00)$ in settings 8–10 respectively. The value of the slope coefficient of the continuous covariate (β_{22}) was then chosen so that the marginal response probability:

$$\Pr(Y_2 = 1) = \frac{p}{[4 - (-4)] \times \beta_{22}} [e^{\beta_{20} + \beta_{21} \times (1) + \beta_{22} \times (4)} - e^{\beta_{20} + \beta_{21} \times (1) + \beta_{22} \times (-4)}] \\ + \frac{1 - p}{[2 - (-6)] \times \beta_{22}} [e^{\beta_{20} + \beta_{21} \times (0) + \beta_{22} \times (2)} - e^{\beta_{20} + \beta_{21} \times (0) + \beta_{22} \times (-6)}]$$

was as large as possible whilst the maximum log binomial probability:

$$\Pr(Y_2 = 1 \mid d = 1, u = 4) = e^{\beta_{20} + \beta_{21} \times (1) + \beta_{22} \times (4)}$$

was at least 90 percent. In the model for Y_1 , we set $\Pr(Y_1 = 1 \mid d = 0, u = -6) = 0.5$, 0.4 and 0.3 and $\Pr(Y_1 = 1 \mid d = 0, u = 2) = 0.12$, 0.15 and 0.18 in settings 8–10 respectively, constrained $\Pr(Y_1 = 1 \mid d = 1, u = 4) = 0.5 \times [1 - \Pr(Y_2 = 1 \mid d = 1, u = 4)]$, and solved for values of the parameters ($\beta_{10}, \beta_{11}, \beta_{12}$) that satisfied these conditions. These designs are summarised in Table 4.

The results for the dichotomous and continuous covariates are shown in Table 5. In the admissible samples, the highest values of average percent relative bias occurred for the coefficient of the dichotomous covariate in the first outcome category (β_{11}) of setting 8. The high values of bias occurred in occasional data replications with a low proportion of $d = 1$ values among observations with $y_1 = 1$. This produced a right-skewed distribution of bias in the estimates of β_{11} , with some very large values particularly for the log multinomial model. The median bias for this model was only 1.33, however. Otherwise bias for the log multinomial estimates did not exceed 5 percent and was generally less than that of the corresponding estimates from the alternative estimators. The log multinomial model estimator also had an advantage in mean squared error. Coverage straddled the 95 percent target and did not favour any model.

The inadmissible samples were produced by inflated estimates of the slope coefficient of the continuous covariate in the second outcome category (β_{22}). The bias and mean squared error of the estimates of that parameter were very high in consequence, and confidence interval coverage was substantially less than the target. This was particularly the case for the log multinomial estimator. For the other slope coefficients, the log multinomial model had slightly superior performance than the alternative models.

5 Example of Fitting a Log Multinomial Model to Real Data

As an example, we use data from a follow-up of participants in the 1985 Australian Schools Health and Fitness Survey. In that survey, an extensive battery of measurements of cardiovascular health, physical activity and fitness was conducted on a nationally representative sample of 7–15 year old

Table 5 Results of estimating the coefficients of the dichotomous ($k = 1$) and continuous ($k = 2$) covariates in settings 8–10.

Setting	Outcome category	Covariate	LnMM*			EMLM*			Poisson*		
	$j = 1, 2$	$k = 1, 2$	Bias †	MSE ‡	Cover §	Bias †	MSE ‡	Cover §	Bias †	MSE ‡	Cover §
Admissible solutions ¶											
8	1	1	8.23	33.09	95.7	6.29	24.73	95.7	6.67	24.28	95.9
	1	2	−0.57	0.25	95.4	−1.83	0.27	95.2	−1.34	0.26	95.3
	2	1	−1.07	4.48	94.6	−3.39	4.84	94.8	−2.52	4.66	95.0
	2	2	−3.83	0.17	96.5	−3.77	0.20	96.4	−5.15	0.19	96.4
9	1	1	3.22	16.93	95.2	2.68	16.96	95.2	2.79	16.89	95.3
	1	2	0.11	0.30	94.8	−2.80	0.33	95.0	−1.74	0.31	95.0
	2	1	1.49	3.51	94.8	0.73	3.60	94.7	1.04	3.55	94.7
	2	2	−3.82	0.10	96.0	−7.08	0.14	96.2	−6.05	0.12	96.1
10	1	1	1.94	17.03	95.1	2.38	17.03	95.0	2.21	16.98	95.1
	1	2	−0.68	0.40	95.1	−6.89	0.43	95.1	−4.45	0.42	95.2
	2	1	0.84	3.31	94.9	0.75	3.35	94.7	0.80	3.33	94.7
	2	2	−3.61	0.06	96.3	−7.09	0.09	96.1	−5.93	0.08	96.2
Inadmissible solutions ¶											
8	1	1	11.44	20.49	96.0	13.80	21.25	96.0	13.40	21.39	95.6
	1	2	3.27	0.25	95.1	5.72	0.29	94.6	4.28	0.27	94.8
	2	1	7.27	3.89	94.0	8.88	4.47	94.9	7.43	4.30	95.1
	2	2	18.13	0.28	88.2	16.07	0.26	92.5	14.76	0.23	92.2
9	1	1	3.66	16.83	95.8	4.96	17.02	96.0	4.76	17.09	95.8
	1	2	4.39	0.30	94.7	10.70	0.34	94.3	8.00	0.32	94.2
	2	1	−2.18	3.16	95.4	0.73	3.45	95.0	−0.24	3.38	95.2
	2	2	21.28	0.19	84.9	20.41	0.20	91.0	18.97	0.17	90.7
10	1	1	1.35	17.27	94.7	0.60	17.19	95.1	0.72	17.30	95.2
	1	2	11.14	0.41	94.7	26.89	0.46	93.9	20.62	0.43	93.9
	2	1	−2.77	3.30	95.0	−0.55	3.36	94.7	−1.16	3.37	94.9
	2	2	25.29	0.13	79.8	24.48	0.14	89.8	22.79	0.12	88.9

* LnMM = log multinomial regression model, EMLM = expanded multinomial logistic regression model, Poisson = separate Poisson regression models.

† Average percent relative bias of the slope estimator.

‡ Average percent mean squared error.

§ Percentage of 95 percent confidence intervals covering the true value of the slope coefficient.

¶ Admissible samples are converged solutions that produced admissible probabilities for the log multinomial model. Inadmissible samples are converged solutions that did not.

Australian schoolchildren. Those aged 9–15 years ($n = 6414$) completed a questionnaire that included questions on attitudes to tobacco smoking. One of those questions required respondents to indicate how important they regarded being a non-smoker. Twenty years later in 2004–05, those remaining in follow-up ($n = 3662$) reported whether they had taken up daily smoking. To illustrate the methodology with a sample of around $n = 500$, we restricted the analyses to male participants at follow-up who were aged 10 or 11 years in 1985. This produced the summary data in Table 6:

Table 6 Childhood attitudes towards smoking and adult smoking status for the sub-sample ($n = 501$) of 10–11 year old male subjects in the Australian Schools Health and Fitness Survey 1985 who participated at follow-up in 2004–05.

Importance of being a non-smoker in 1985	Smoking status in 2004–05			Total
	Never smoked	Former smoker	Current smoker	
Not very important	24	11	32	67
Very important	257	65	112	434
	281	76	144	501

The purpose of the analysis was to obtain estimates of the relative risk of each category of adult smoking status in 2004–05 for not regarding non-smoking as being very important in 1985, when subjects were 10 or 11 years of age. Further, these estimates were to be adjusted for exact age in 2004–05 (range 27.3 years to 32.2 years). The outcome variable for the analysis of these data was *daily_smoking* (Y) with its three categories coded as $Y = 0$ (never smoked), $Y = 1$ (former smoker) and $Y = 2$ (current smoker). The covariates were a binary (0/1) indicator for not regarding being a non-smoker as very important (*ambivalent*), and a continuous predictor for age at follow-up (*age*). The values of *age* were centred by subtracting from them the mean age for the sub-sample (28.8 years).

The results of using Stata's *ml* command to fit a log multinomial regression model to these data are shown in Table 7. The intercept (*_cons*) and slope (*ambivalent*) in the model without adjustment for age correctly estimate the log baseline risk and log relative risk respectively, and yield values of risk and relative risk that match those able to be calculated from the summary data in Table 6. The standard errors in Table 7 are robust estimates calculated from the observed information using the information sandwich estimator (4). When estimated from the Hessian (3), the estimates are $\widehat{SE}(\hat{\beta}_{10}) = 0.114370$ and $\widehat{SE}(\hat{\beta}_{11}) = 0.298436$ for former smoking, and $\widehat{SE}(\hat{\beta}_{20}) = 0.081391$ and $\widehat{SE}(\hat{\beta}_{21}) = 0.151489$ for current smoking. These match exactly those able to be calculated using standard formula for the standard error of the logarithm of risk and of relative risk (Katz et al., 1978). For a model with a single dichotomised covariate model, the standard errors are unchanged if the expected information matrix is used in place of the observed information, and the coefficients and standard errors are identical to those produced by fitting log binomial models to the outcome categories separately (data not shown).

To adjust for age at follow-up, the continuous covariate *age* was added to the log multinomial model. The results are shown in the lower panel of Table 7. Doing so slightly reduced the estimated log relative risk of being a former smoker for having an ambivalent attitude to smoking in childhood, and slightly increased the estimated log relative risk of being a current smoker. This was as expected because the mean age at follow-up of subjects with an ambivalent attitude (28.9 years) was higher than the mean age of those who regarded non-smoking as very important (28.8 years), but only slightly.

6 Discussion

In this paper, we proposed an extension of the log binomial model to the multinomial outcome setting and investigated the feasibility of fitting the model. The performance of the model was assessed by a simulation-based comparison of the estimators of its coefficients and standard errors with those of two alternative approaches. The first of these was an extension to multinomial outcomes of the expanded

Table 7 Results of fitting a log multinomial regression model to the smoking status and attitudes data for the sub-sample ($n = 501$) of 10–11 year old male subjects in the Australian Schools Health and Fitness Survey 1985 who participated at follow-up in 2004–05.

	Former smoker					Current smoker				
	Coeff.	Std.Err	z	$P > z $	95% Conf.Int.	Coeff.	Std.Err	z	$P > z $	95% Conf.Int.
	Unadjusted									
<i>ambivalent</i>	0.092	0.299	0.31	0.758	−0.494, 0.677	0.616	0.152	4.06	0.000	0.318, 0.913
<i>_cons</i>	−1.899	0.114	−16.58	0.000	−2.123, −1.674	−1.355	0.081	−16.63	0.000	−1.514, −1.195
	Adjusted for age									
<i>ambivalent</i>	0.085	0.298	0.28	0.771	−0.499, 0.668	0.618	0.152	4.07	0.000	0.320, 0.916
<i>age</i>	0.389	0.164	2.38	0.017	0.069, 0.710	−0.044	0.114	0.38	0.701	−0.270, 0.181
<i>_cons</i>	−1.934	0.123	−15.66	0.000	−2.176, −1.692	−1.355	0.082	16.59	0.000	−1.516, −1.195

data logistic regression model proposed by Schouten et al. (1993) for binary data. The second involved separate fits of a Poisson regression model to all but one of the outcome categories. Finally, we investigated a real data example to verify the veracity of the model estimates.

While currently there is no standard software to fit the model, specifying the multinomial likelihood and fitting the model by modified Newton–Raphson optimisation of the log likelihood was straightforward with the *ml* command in the Stata software package (Gould, Pitbaldo and Sribney, 2003). The real data example demonstrated that the log multinomial model produced coefficient and standard error estimates that matched exactly those calculated from the summary data using standard formulae. Furthermore, as should be the case in the case of a single dichotomous covariate, they matched exactly those from fitting binary log binomial models to the relevant outcome categories. The results of adjusting for a continuous covariate were as expected.

The simulations of log multinomial data demonstrated the occurrence of inadmissible solutions with this model, but the two alternative methods produced out-of-bounds solutions with similar frequency. Even when a log multinomial model was successfully fitted, the alternative models fitted to the same data produced inadmissible values for the log multinomial model on up to 28 percent of fits by the expanded data multinomial logistic regression model and up to 24 percent of fits by the separate Poisson regression models.

Of the three estimators, the log multinomial estimator provided slope estimates with the smallest average percent relative bias and smallest average percent mean squared error in the majority of admissible samples from the data settings we investigated. The expanded data multinomial logistic estimator had the poorest performance in these respects, but the differences were really quite small. Thus whilst the log multinomial regression model had slightly superior performance on these criteria, the expanded data multinomial logistic estimator and – in particular – the separate Poisson regression estimator each provides a close approximation to it. Only in one setting did average bias of a coefficient estimated by a log multinomial regression model exceed 5 percent, but the distribution of bias was right-skewed and median bias was just one percent. The large values in its right-skewed distribution arose from appropriate fits of the log multinomial regression model to datasets that were extreme but reflected the inherent variability in the distribution of log multinomial data for that setting.

In the inadmissible samples, the bias and mean squared error of one or more of the slope estimates was unacceptably high in three of the settings we investigated. These were settings with models having two covariates, which were the most realistic we investigated. This cautions against drawing inferences from models with inadmissible solutions.

For each estimator, the confidence interval coverage of the slope parameters straddled 95 percent in the admissible samples. The variance estimator was an information sandwich estimator based on the observed information matrix. We could have used instead the non-robust version of it, or substituted the expected value of the observed information matrix when calculating either. In comparisons of results for these four variance estimators (data not reported), coverage was more often closer to 95 percent when calculated with the robust estimates than when calculated with the non-robust estimates, and when based on the observed information matrix than when based on the expected information matrix. Whilst these differences were slight, they suggest that – at least in samples of this size and for simple models of the types we considered – the robust estimates based on the observed information matrix have superior sampling properties.

There are several limitations of our research that are worth keeping in mind. Firstly, we focused on outcomes with three categories, but the model extends readily to nominal outcome data of higher dimensions. Secondly, our results are based primarily on data simulations, and we did not simulate multivariable models that are as complicated as those typically encountered in practice. Thirdly, we used results based on maximum likelihood theory to compute Wald-type confidence intervals. The log multinomial model places a theoretical constraint on estimators that the linear predictor must be negative for all subjects in the sample, and hence assumptions about the parameter space required for distributional theory do not hold. Finally, we have not addressed the issue of data having a solution that lies on the boundary of the permissible parameter space. In the binary context, Deddens and

colleagues (Deddens, Petersen and Lei, 2003; Deddens and Petersen, 2004) suggest estimating the model on an expanded dataset formed by combining $(c - 1)$ copies of the original data with one version of the data that has the values reversed for the binary outcome indicator. They refer to this as the COPY algorithm. For their data example, our method produces a convergent, admissible solution that is a close approximation to the maximum likelihood solution (Blizzard and Hosmer, 2007). Lumley, Kronmal and Ma (2006) reviewed this and other approaches to the problem, and recommended an alternative solution. We verified that the COPY method can be extended with the desired result to simple examples of trinomial outcome data but did not investigate further.

In summary, we feel that the log multinomial model does offer a practical solution to the problem of obtaining adjusted estimates of the risk ratio in the multinomial setting. Particularly when the sum of the fitted probabilities is high, however, the model must be used with some care and attention to detail.

Acknowledgements The authors thank the referees whose comments and suggestions lead to a greatly improved paper.

References

- Barros, A. J. and Hirakata V. N. (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology* **3**, 21.
- Blizzard, L. and Hosmer D. W. (2006). Parameter estimation and goodness-of-fit in log binomial regression. *Biometrical Journal* **48**, 5–22.
- Blizzard, L. and Hosmer D. W. (2007). Author's Reply. *Biometrical Journal* **49**, 330–331.
- Carter, R. E., Lipstiz, S. R., and Tilley, B. C. (2005). Quasi-likelihood estimation for relative risk regression models. *Biostatistics* **6**, 39–44.
- Deddens, J. A., Petersen, M. R., and Lei, X. (2003). Estimation of prevalence ratios when PROC GENMOD does not converge. Proceedings of the 28th Annual SAS Users Group International Conference, paper 270–28, Cary NC, SAS Institute Inc. <http://www2.sas.com/proceedings/sugi28270-28.pdf>
- Deddens, J. A. and Petersen, M. R. (2004). Re: Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology* **159**, 213–214.
- Gould W., Pitbaldo J., and Sribney W. (2003). *Maximum Likelihood Estimation with Stata*: Second Edition. Stata Corporation, Texas.
- Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analysis. *American Journal of Epidemiology* **125**, 761–768.
- Greenland, S. (2004). Model-based estimation of relative risks and other epidemiologic measures of common outcomes and in case-control studies. *American Journal of Epidemiology* **160**, 301–305.
- Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*: Second Edition. Wiley, New York.
- Katz, D., Baptista, J., Azen, S. P., and Pike, M. C. (1978). Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics* **34**, 469–474.
- Katz, K. A. (2006). The (relative) risks of using odds ratios. *Archives of Dermatology* **142**, 761–764.
- Lee, J. (1994). Odds ratio or relative risk for cross-sectional data? *International Journal of Epidemiology* **23**, 201–203.
- Lumley, T., Kronmal, R., and Ma, S. (2006). Relative risk regression in medical research: models, contrasts, estimators, and algorithms. *UW Biostatistics Working Paper Series* 293. <http://www.bepress/uwbiostat/paper293>
- McCullough, P. and Nelder, J. (1989). *Generalized Linear Models*: Second Edition. Chapman Hall: New York.
- McNutt, L.-A., Wu, C., Xue, X., and Hafner, J. P. (2003). Estimating relative risk in cohort studies and clinical trials of common events. *American Journal of Epidemiology* **157**, 940–943.
- Schouten, E. G., Dekker, J. M., Kok, F. J., le Cessie, S., van Houwelingen, H. C., and Pool, J. (1993). Risk ratio and rate ratio estimation in case cohort designs: hypertension and cardiovascular mortality. *Statistics in Medicine* **12**, 1733–1745.
- Sinclair, J. C. and Bracken, M. B. (1994). Clinically useful measures of effect in binary analyses of randomized trials. *Journal of Clinical Epidemiology* **47**, 881–889.
- Skov, T. Deddens, J., Petersen, M. R., and Endahl, L. (1998). Prevalence proportion ratios: estimation and hypothesis testing. *International Journal of Epidemiology* **27**, 91–95.

- Stata Corp. (2005). *Stata Statistical Software: Release 9.0*. Stata Corporation: College Station Texas.
- Spiegelman D. and Hertzmark E. (2005). Easy SAS calculations for risk or prevalence ratios and differences. *American Journal of Epidemiology* **162**, 199–200.
- Zochetti, C., Consinni, D., and Bertazzi, P. A. (1995). Estimation of prevalence rate ratios from cross-sectional data. *International Journal of Epidemiology* **24**, 1064–1065.
- Wacholder, S. (1986). Binomial regression in GLIM: estimating risk ratios and risk differences. *American Journal of Epidemiology* **123**, 174–184.
- Zou, G. (2004). A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* **154**, 702–706.