ELSEVIER
Computational Statistics & Data Analysis 39 (2002) 57–74
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

A modified score function estimator for multinomial logistic regression in small samples

Shelley B. Bull[a,*], Carmen Mak[b], Celia M.T. Greenwood[c]

[a] *Samuel Lunenfeld Research Institute and Department of Public Health Sciences, University of Toronto, Toronto, Ont. Canada*
[b] *Schering Plough Research Institute, Department of Statistics, Kenilworth, NJ, USA*
[c] *Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada*

Received 1 January 2001; received in revised form 1 June 2001; accepted 1 June 2001

Abstract

Logistic regression modelling of mixed binary and continuous covariates is common in practice, but conventional estimation methods may not be feasible or appropriate for small samples. It is well known that the usual maximum likelihood estimates (MLEs) of the log-odds-ratio parameters are biased in finite samples, and there is a non-zero probability that an MLE is infinite, i.e., does not exist. In this paper, we extend the approach proposed by Firth (Biometrika 80 (1993) 27) for bias reduction of MLEs in exponential family models to the multinomial logistic regression model, and consider general regression covariate types. The method is based on a suitable modification of the score function that removes first order bias. We apply the method in the analysis of two datasets: one is a study of disease prognosis and the other is a disease prevention trial. In a series of simulation studies in small samples, the modified-score estimates for binomial and trinomial logistic regressions had mean bias closer to zero and smaller mean squared error than other approaches. The modified-score estimates have properties that make them attractive for routine application in logistic regressions of binary and continuous covariates, including the advantage that they can be obtained in samples in which the MLEs are infinite.

Keywords: Asymptotic bias; Bayesian estimates; Bias reduction; Continuous covariate; Infinite estimates; Jeffreys' prior; Odds ratio; Polychotomous logistic regression; Polytomous logistic regression; Sparse data

* Corresponding author. Samuel Lunenfeld Research Institute, Suite 850, 600 University Avenue, Toronto, Ont. Canada M5G 1X5. Tel.: +416-586-8245; fax: +416-586-8404.
E-mail address: bull@mshri.on.ca (S.B. Bull).

0167-9473/02/$ - see front matter
PII: S 0167-9473(01)00048-2

1. Introduction

Methods for logistic regression modelling of nominal categorical responses based on the multinomial logistic likelihood are now generally available in standard statistical packages, and have been applied in the analysis of case-control studies with multiple case or multiple control groups, and in randomized trials and cross-sectional surveys with categorical responses. One of the concerns of investigators is the valid estimation of model parameters in the finite sample sizes encountered in practice. In finite samples, the usual maximum likelihood estimates (MLEs) of the log odds ratios are biased, and the bias increases as the ratio of the number of observations to the number of parameters (n-to-p ratio) decreases (Cordeiro and McCullagh, 1991; Bull et al., 1997). This is of particular concern when there are several response categories and multiple covariates because the number of parameters can become large.

We consider an alternative estimation method for small samples based on a modification of the score function that removes first order bias and is equivalent to penalizing the likelihood by Jeffreys' prior (Firth, 1993). We extend the modified score function method to the multinomial logistic regression model with nominal response categories, and compare the modified estimates to the usual MLEs and to the MLEs corrected by an estimate of the asymptotic bias. As the sample size increases, the modified-score estimates become equivalent to the usual MLEs. As systematic small sample comparisons of this approach have not been reported previously for binomial or multinomial logistic regression models, we also present a Monte Carlo simulation study in which we compare the mean bias and mean squared error (MSE) of the modified estimates to the MLEs, and to the MLEs corrected by the estimated asymptotic bias. Considering the same series of logistic regression models studied previously (Bull et al., 1997), we find that the modified-score estimates are competitive and often superior to the other approaches.

2. Methods for small-sample analysis

The small-sample properties of the logistic regression MLEs can be improved by the general approach of Cox and Snell (1968) which uses higher order terms in a Taylor series expansion of the log-likelihood to approximate the asymptotic bias and obtain bias-corrected MLEs (Anderson and Richardson, 1979; Schaefer, 1983; Copas, 1988; Cordeiro and McCullagh, 1991; Bull et al., 1997). When the magnitude of the linear predictor is small, Cordeiro and McCullagh (1991) showed that the effect of bias correction in generalized linear models is to shrink the MLEs toward the origin by a factor that depends on the ratio of the number of parameters to the sample size and on the magnitude of the underlying log-odds-ratio parameter. The asymptotic bias correction method reduces bias and MSE (Anderson and Richardson, 1979; Schaefer, 1983), but in small samples tends to correct beyond the true value (Bull et al., 1997). Bias reduced estimates can also be obtained by jackknife methods, with or without full iteration, but the fully iterated estimates, particularly, are also over correct (Farewell, 1978; Bull et al., 1994, 1997). Both methods reduce the bias to order n^{-1}, approximately, but do not remove it completely, and both methods require that finite MLEs can first be obtained.

When data are sparse or unbalanced, estimates of logistic regression parameters from a conditional likelihood may have better properties than those from the usual unconditional likelihood (Hauck, 1984, for example). The analysis of stratified 2×2 tables arises as a special case in which the strata indicators are nuisance parameters. In this approach, the conditional likelihood for the common log-odds-ratio parameter is obtained by conditioning on the sufficient statistics of the strata parameters. However, the conditional logistic likelihood is not immune to the problems of finite sample bias and infinite estimates either. Greenland et al. (2000) emphasize that statistical bias may produce unduly large estimates when data are sparse or many variables are controlled.

More generally when there are multiple covariates of interest, there may be no natural nuisance parameters per se. In this case, Mehta and Patel (1995) note that exact conditional inferences can be obtained for each parameter in turn by treating the other parameters in the regression as nuisance parameters and conditioning jointly on their sufficient statistics. For general binomial logistic regression models, exact methods based on the conditional likelihood are available in the software package LogXact (Cytel Software Corporation, 1992). Because this conditional likelihood involves conditioning on the observed covariate values, a problem with overconditioning can arise when one or more of the covariates is continuous; in this case, support for the conditional distribution of the sufficient statistic for the parameter of interest can become extremely discrete or even degenerate (LogXact Software Manual, Cytel Software Corporation, 1992). To avoid overconditioning, continuous covariates can be grouped into a smaller number of levels, but this is not an ideal solution for regressions that include several continuous covariates, and may compromise efficiency or the ability to control for continuous confounders (Greenland, 1995). Theory and methods have also been developed for multinomial responses with nominal and ordinal categories (Zelen, 1991; Hirji, 1992), but are not yet generally available in software.

Several authors have evaluated alternative estimation methods for conditional and unconditional logistic regression motivated by Bayesian inference, and have examined normal, uniform, or Jeffreys priors (Rubin and Schenker, 1987; Duffy and Santner, 1989; Santner and Duffy, 1989; see Section 5.4A; Ibrahim and Laud, 1991; Firth, 1992a,b, 1993; Greenland et al., 2000; Greenland, 2000). Use of these prior distributions tends to shrink the estimates toward the origin, and thereby reduces bias and mean squared error. In the context of a matched pair study of a dichotomous exposure, Greenland (2000) found that the expected bias and squared error of several Bayes odds-ratio estimators were smaller than for the uncorrected and bias-corrected conditional MLEs. Rubin and Schenker (1987), Greenland et al. (2000), and Firth (1993) note the equivalence between the well-known Haldane correction, which adds $\frac{1}{2}$ to each cell, and a Bayesian estimator based on Jeffreys' prior. In exponential family models with canonical parameterization, Firth shows that modifying the score function to remove first order bias is equivalent to penalizing the likelihood by the

Jeffreys prior and removes the order n^{-1} bias of the MLEs. In an extension of this idea in which the likelihood score for right censored data having an exponential distribution is modified similarly, Pettitt et al. (1998) found that the modified estimates have smaller bias and variance than the MLEs and good bias properties irrespective of the existence or otherwise of the MLEs. In the next section, we generalize this Bayesian-motivated approach to multinomial logistic regression with general covariate types.

3. Estimation and computation

3.1. Usual maximum likelihood estimation with bias correction

We consider a multicategory outcome y that is a multinomial variable with $J+1$ categories. For each category j $(j=1,\ldots,J)$, there is a regression function in which the log odds of response in category j, relative to category 0, is a linear function of regression parameters and a vector $\mathbf{x}$ of p covariates (including a constant): $\log\{prob(y=j\mid\mathbf{x})/prob(y=0\mid\mathbf{x})\}=\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x}$. We let $\mathbf{y}_i$ be a $J\times 1$ vector of indicators for the observed response category for observation i, with the corresponding $J\times 1$ vector of probabilities $\theta_i=(\theta_{i1},\ldots,\theta_{iJ})^{\mathrm{T}}$. We estimate the vector of uncorrected MLEs, $\hat{B}=vec[(\hat{\boldsymbol{\beta}}_1\ldots\hat{\boldsymbol{\beta}}_J)^{\mathrm{T}}]$, from observations $(\mathbf{y}_i,\mathbf{x}_i)$, $i=1,\ldots,n$, by solving the score equations of the log-likelihood $l(B)$.

The score function is $U(B)=X^{\mathrm{T}}(Y-\boldsymbol{\Theta})$, with $X^{\mathrm{T}}=X_D^{\mathrm{T}}\otimes I_J$, $X_D^{\mathrm{T}}=(\mathbf{x}_1\mid\mathbf{x}_2\mid\ldots\mid\mathbf{x}_n)$, $Y=(\mathbf{y}_1|\mathbf{y}_2|\ldots|\mathbf{y}_n)$, and $\boldsymbol{\Theta}=(\theta_1\mid\theta_2|\ldots|\theta_n)$; $\otimes$ is the Kronecker product matrix operator. The $pJ\times pJ$ Fisher information matrix is $A=(X^{\mathrm{T}}MX)$, and M is an $nJ\times nJ$ block diagonal matrix with n $J\times J$ blocks $M_i=\{m_{ijk}\}$, $m_{ijj}=\theta_{ij}\,(1-\theta_{ij})$ for $j=k$ and $m_{ijk}=-\theta_{ij}\theta_{ik}$ otherwise. Using t to denote the iteration number, the iterative updating equation used to obtain the MLEs is

$$B_{(t+1)}=B_{(t)}+A_{(t)}^{-1}U(B_{(t)}) \tag{3.1}$$

with $U(B_{(t)})=X^{\mathrm{T}}\,(Y-\boldsymbol{\Theta}_{(t)})$, and $\boldsymbol{\Theta}_{(t)}$ and $A_{(t)}$ evaluated at $B_{(t)}$.

Asymptotically, the MLEs are normally distributed around the true parameter with variance given by the inverse of the Fisher information matrix. Standard errors are conventionally obtained by evaluating A^{-1} at the MLEs. In finite samples, the quadratic approximation to the log likelihood may not apply, and Wald test statistics and confidence intervals based on the large sample standard errors may have poor properties when the parameter is far from zero (Hauck and Donner, 1977; Jennings, 1986). When the normality of the MLE is in doubt, confidence intervals based on likelihood ratio tests or on score tests have better properties (Alho, 1992 and references therein).

The leading term in the asymptotic bias of the MLEs:

$$b(B)=-\frac{1}{2}\,A^{-1}\{X^{\mathrm{T}}Q(X\otimes X)\,vec\,(A^{-1})\} \tag{3.2}$$

is obtained from the Taylor series expansion of the log likelihood of B; $X^{\mathrm{T}}Q\,(X\otimes X)$ is the matrix of third derivatives with respect to B (Bull et al., 1997). The

$nJ \times (nJ)^2$ matrix Q has a $n \times n^2$ block structure, with $J \times J^2$ submatrices Q_i such that $Q = \sum_i E_i \, Q_i \, (E_i \otimes E_i)^{\mathrm{T}}$, where $E_i = \mathbf{e}_i \otimes I_J$; $\mathbf{e}_i$ is a unit vector of length n, with 1 in position i and 0's elsewhere. The elements of the submatrix Q_i depend on the probabilities θ_{ij}, $j = 1, 2, \ldots, J$ with $Q_i = \sum_{jkl} q_{ijkl} \mathbf{i}_j (\mathbf{i}_k \otimes \mathbf{i}_l)^{\mathrm{T}}$, where $q_{ijkl} = \theta_{ij}(1 - \theta_{ij})\,(1 - 2\theta_{ij})$ for $j = k = l$, $2\theta_{ij}\theta_{ik}\theta_{il}$ for $j \neq k \neq l$, $-\theta_{ij}\,(1 - 2\theta_{ij})\,\theta_{il}$ for $j = k \neq l$, and $-\theta_{ij}\theta_{ik}\,(1 - 2\theta_{il})$ for $j \neq k$ and $l = j$ or $l = k$; $\mathbf{i}_j$ is a unit vector of length J, with 1 in position j and 0's elsewhere. As originally described by Cox and Snell (1968), the bias corrected estimates (BCMLE) from the asymptotic expansion are

$$B_{\mathrm{BC}} = \hat{B} - b(\hat{B}) \tag{3.3}$$

with A and Q evaluated at the uncorrected MLEs.

In finite samples, there is a non-zero probability that an MLE is infinite, i.e., does not exist. Existence problems can occur when the data are sparse or when there are large covariate effects. This situation is characterized by separation in the sample space among the groups that correspond to the categories of the outcome variable, and failure of one or more of the elements of B to converge. Albert and Anderson (1984), Santner and Duffy (1986), and Lesaffre and Albert (1989) discuss the definition and identification of infinite estimates for multiple covariates. It is not unusual for separation to occur in small and moderate-sized datasets, especially in multinomial logistic regression. To detect separations, we applied the same general algorithm that we adapted from others and used in previous studies (Albert and Harris, 1984; Lesaffre and Albert, 1989; Bull et al., 1997). When the estimate of a parameter is infinite, the corresponding bias correction cannot be calculated.

3.2. Estimation using a modified score function

The modified score function proposed by Firth for the binomial logistic model extends directly to the multinomial model as $U^*(B) = U(B) - Ab(B)$, where A is the Fisher information for the MLEs and $b(B)$ is their asymptotic bias defined in (3.2) above. The introduction of bias into the score function removes the leading term in the asymptotic bias of the MLE. As described by Firth (1993), the solution of $U^* = 0$ locates a stationary point of $l^*(B) = l(B) + \frac{1}{2}\log|A|$ which is equivalent to the penalized likelihood function $L^*(B) = L(B)|A|^{1/2}$ with the Jeffreys' invariant prior as the penalty function. Use of the Jeffreys' prior shrinks estimates toward the point $\theta_{ij} = 1/(J + 1)$ that maximizes the determinant, and corresponds to $\boldsymbol{\beta}_j = 0$. Taking the derivative of the log-likelihood penalty with respect to B yields the modification to the score function. The resulting estimate B_{MS} agrees with B_{BC} to second order. The arguments for the existence of estimates in the binomial model given by Firth extend to the multinomial model as follows. Provided X_D is of full rank, $\log|A|$ is strictly concave and unbounded below as B goes to infinity in any direction. Then, combined with the condition that the log likelihood $l(B)$ is strictly concave and bounded above, the maximum penalized likelihood estimates (MPLEs) exist in any finite sample and are unique.

The penalty function can be applied using iterative adjustments; the modified iterative updating equations are

$$B^*_{(t+1)} = B^*_{(t)} + A^{-1}_{(t)} U^*(B^*_{(t)})$$
$$= B^*_{(t)} + A^{-1}_{(t)} \{X^{\mathrm{T}}(Y - \mathbf{\Theta}_{(t)}) + \frac{1}{2} X^{\mathrm{T}} Q_{(t)} (X \otimes X)\, vec(A^{-1}_{(t)})\} \qquad (3.4)$$

with $\mathbf{\Theta}_{(t)}$, $A_{(t)}$ and $Q_{(t)}$ evaluated at $B^*_{(t)}$. This can be rewritten as

$$B^*_{(t+1)} = B^*_{(t)} - b(B^*_{(t)}) + A^{-1}_{(t)} U(B^*_{(t)}). \qquad (3.5)$$

Thus, the modification operates by applying the asymptotic bias corrections at each step in the iterative process; this prevents the estimates from going off to infinity and failing to converge when there is separation in the data. In contrast, the conventional asymptotic method applies the bias correction at the end, after the process has converged to the MLEs, and thus it is not possible to apply bias corrections when separation occurs.

Firth (1992a,b, 1993) discusses practical aspects of this approach for the binomial logistic model, including implementation in a reweighted least squares framework and specification of starting values, and demonstrates that the modified procedure converges at a linear rate. In the multinomial model, however, a reweighted least squares approach is less easily implemented because the weight matrix M is block diagonal rather than simply diagonal. We implemented a modified Fisher scoring algorithm based on (3.5) using the matrix programming language GAUSS (Aptech Systems, 1990). This is not a true scoring or Newton-type algorithm because it updates with the inverse of A, the Fisher information for the MLE's, rather than the information for the MPLEs, A^*, which includes an additional term corresponding to the second derivatives of the penalty $\frac{1}{2}\log|A|$. As a result, convergence using the modified scoring algorithm for the MPLEs is slower than for a scoring algorithm based on A^*, which converges at a quadratic rate. For well behaved and larger datasets, this usually meant no more than 2 or 3 steps beyond that required for the MLEs. We found starting values of $\boldsymbol{\beta}_j = 0$ to be generally satisfactory. For smaller datasets, i.e., less than 50 observations, and especially datasets in which there were infinite MLEs, convergence was slower and could take up to 35 or 40 iterations. In datasets with separations, starting values other than 0 could lead to divergence of the process.

The standard error estimates of the MPLEs obtained from the Fisher information matrix evaluated at the MPLEs, which can be referred to as plug-in estimates, are smaller than those for the MLEs because the MPLEs are generally smaller than the MLEs and the standard errors tend to be proportional to the magnitude of the regression estimates. Although, as Firth (1993) notes, the first order asymptotic covariance matrix of the MPLEs is the same as that of the MLEs, construction of symmetric confidence intervals based on the MPLEs and the plug-in standard errors may be ill advised because the small sample situations in which MPLEs are needed will also be those in which the quadratic approximation of the log likelihood does not apply (Jennings, 1986). In their application of modified-score estimates, Pettitt et al. (1998) suggest a parametric bootstrap method for determination of standard errors and confidence intervals. We prefer to construct asymmetric confidence intervals based on the profile likelihood using methods analogous to those implemented in SAS PROC LOGISTIC for the MLEs (Venzon and Moolgavkar, 1988; SAS/STAT, 1996).

4. Applications of small-sample estimation

The first application is a study of clinical factors that relate to the presence or absence of nodal involvement in patients with prostate cancer; 20 of the 53 patients have nodal involvement (Brown, 1980; Cox and Snell, 1989). Table 1 presents the estimates for a logistic regression model with six covariates: four binary indicators (X-ray, stage, grade, interaction of stage and grade) and two continuous variables (acid, age). In this dataset, the ratio of the number of observations to the number of parameters is less than 10, and previous studies indicate that the MLEs are likely to be inflated in magnitude, especially for the covariates with strong effects (Bull et al., 1997). The bias reduced estimates are 18–24 percent smaller in magnitude than the MLEs, with the MPLEs approximately 20 percent smaller and the BCMLEs 23 percent smaller. As expected, the corresponding standard errors, obtained by evaluating the Fisher information matrix at the bias reduced estimate values, are also smaller than those for the MLEs. Thus, although the MLEs all exist, the magnitude of the prognostic covariate effects are likely being overstated as a result of finite sample bias in the MLEs. This sample size is smaller than desirable for a definitive prognostic study, but use of these MLEs to design a larger study would likely yield an underestimate of the true sample size requirements.

A second application is from a multicentre trial of a population intervention to prevent post-transfusion hepatitis (PTH), which was classified into two types: hepatitis C and non-A non-B non-C hepatitis (Blajchman et al., 1995). The preventive intervention (Treatment factor) under evaluation was the screening of donor units for two surrogate markers of non-A non-B hepatitis infection. Blood transfusion recipients were randomized to receive units from one of two sources: from the general blood supply or from a supply that had excluded units that were positive for the surrogate

Table 1
Nodal involvement in prostate cancer (two response categories, six covariates, $n = 53$). Maximum likelihood and alternative estimates with asymptotic standard errors

Covariate	MLE[a]	(se)	BCMLE[b]	(se)	MPLE[c]	(se)
Intercept	−10.95	(6.65)	−8.62	(5.95)	−8.93	(6.07)
X-ray	2.61	(0.99)	2.02	(0.85)	2.12	(0.87)
Stage	3.65	(1.29)	2.78	(1.08)	2.99	(1.11)
Grade	4.04	(1.55)	3.07	(1.33)	3.24	(1.36)
Stage ∗ Grade	−4.93	(1.96)	−3.73	(1.70)	−3.94	(1.74)
Log (Acid)	3.02	(1.35)	2.35	(1.19)	2.45	(1.21)
Age	−0.09	(0.07)	−0.07	(0.06)	−0.07	(0.06)

[a]MLE: maximum likelihood estimate.
[b]BCMLE: asymptotic bias corrected estimate.
[c]MPLE: modified score function estimate based on penalized likelihood.

Table 2
Hepatitis prevention trial data

	Hepatitis outcome:		
	C	Non-ABC	No disease
Time 1			
Treated	0	2	400
Untreated	5	3	389
Time 2			
Treated	3	10	1896
Untreated	5	11	1864

Table 3
Hepatitis prevention trial (three response categories, three covariates, $n = 4588$). Maximum likelihood and alternative estimates with 95% confidence intervals

Outcome: Covariate	MLE[a]	(95% CI)[b]	MPLE[c]	(95% CI)[b]	CMLE[d]	(95% CI)
Hepatitis C						
Treatment	—[e]	$(-\infty, -0.78)$	-2.43	$(-7.30, -0.24)$	-1.93	$(-\infty, 0.07)$
Time	-1.57	$(-2.85, -0.28)$	-1.57	$(-2.79, -0.34)$	-1.57	$(-3.04, -0.09)$
Treatment × Time	—	$(-0.15, +\infty)$	1.96	$(-0.70, 6.95)$	1.10	$(-1.32, +\infty)$
Non-ABC Hepatitis						
Treatment	-0.43	$(-2.46, 1.37)$	-0.36	$(-2.16, 1.28)$	-0.43	$(-2.92, 1.74)$
Time	-0.27	$(-1.44, 1.22)$	-0.38	$(-1.49, 0.99)$	-0.27	$(-1.56, 1.46)$
Treatment × Time	0.32	$(-1.67, 2.50)$	0.26	$(-1.58, 2.22)$	0.31	$(-2.07, 2.98)$

[a]MLE: maximum likelihood estimate in trinomial regression.
[b]Based on profile likelihood (SAS Proc Logistic).
[c]MPLE: modified score function estimate in trinomial regression.
[d]CMLE: conditional maximum likelihood estimate in two binomial regressions (LogXact).
[e]Infinite parameter estimate.

markers. In addition, while the trial was on-going there was a change in national blood screening policy whereby a new test was introduced to screen all units for hepatitis C antibodies (time factor). This had the effect of decreasing the incidence of hepatitis C. Thus to evaluate whether the intervention was equally effective before and after the change in screening policy, a test for interaction between the treatment and time factors would be of interest.

The disease outcomes are rare, producing empty cells in some subgroups (Table 2). As a result, in the model with an interaction between time and treatment, the usual logistic regression MLEs are infinite for two of the parameters, and corresponding BCMLEs cannot be obtained. The MPLEs, however, can be obtained for all parameters (Table 3). In this simple case of two 2×3 tables, the MPLEs correspond to calculating the usual MLEs after adding $\frac{1}{2}$ to all the cells in the tables, a strategy originally suggested for sparse tables by Haldane (1956). When the covariate has a balanced distribution, as in the randomized treatment assignment, the MPLEs and their standard errors are smaller than the unconditional logistic regression MLEs. However, when the covariate is unbalanced, as in the time factor, the MPLEs can be larger in magnitude than the MLEs, although their standard errors are smaller. The profile likelihood confidence intervals for the infinite MLEs have one infinite limit, indicating that $\beta = \pm\infty$ cannot be ruled out. This corresponds to the possibility that a person with a given covariate value can be affected or unaffected with certainty, which is implausible in most cases, and in this study more likely reflects the low frequency of the hepatitis C outcome. In contrast, the profile likelihood confidence intervals for the MPLEs have finite limits.

In the corresponding conditional logistic model, we obtained the conditional MLEs from LogXact-Turbo (Cytel Software Corporation, 1992) by fitting two separate binomial logistic models, one that compared the recipients with hepatitis C to those without disease and another that compared those with non-ABC hepatitis to those with no disease. Because the model with an interaction is saturated, the exact binomial results are equal to the exact multinomial results, but this will not be true in general. If other covariates were to be modelled, an implementation of exact multinomial regression would be required. Here we see that the CMLEs are similar to the unconditional MLEs when the latter exist, but the confidence intervals are somewhat larger (Table 3). For the parameters in the hepatitis C regression that have infinite MLEs, however, the CMLEs are smaller in magnitude than the MPLEs, and the confidence intervals have one infinite limit. Furthermore, if we wanted to include a continuous covariate related to the risk of PTH, such as age, without having to group it into a smaller number of levels, then the conditional approach might be complicated by over-conditioning.

5. A small sample Monte Carlo simulation study

5.1. Design

The Monte Carlo study evaluated the MPLEs with respect to mean bias and MSE and compared them to the usual MLEs and the BCMLEs in multiple logistic regressions that included both binary and continuous covariates. We also calculated the mean bias and MSE for the MPLEs in all datasets, including those in which one or more of the MLEs did not exist. As in our previous study, we conducted three series of simulations to investigate the effects of sample size $(200, 100, 75, 50)$, type of covariates (binary and normal), number of parameters (three or six), parameter values, and correlation among the covariates (uncorrelated or correlated). We chose regression parameter values to facilitate comparisons of mean bias for specific values of the slope parameters, and to examine the effects of n and $\boldsymbol{\beta}_j$ for fixed J and p, and the effects of J and p for fixed n and $\boldsymbol{\beta}_j$.

We programmed the simulations in the matrix language GAUSS, using the random number generators provided therein. To generate a vector of mixed binary and normal covariates for each observation in a dataset, we first generated variates from a multivariate normal distribution with zero means, unit variances, and a specified correlation matrix, followed by dichotomization at zero to produce one or more binary covariates. We estimated the binary covariate correlations induced by dichotomization from the simulated covariate vectors. We generated the response category by comparing the probabilities calculated from the linear predictor(s), $\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x}_i$, to a uniform random number. To obtain datasets and results directly comparable to those of a previous study of jackknife bias reduction (Bull et al., 1997), we used the same random number seeds as earlier.

One series of simulations evaluated regressions of binomial responses on two covariates, one binary and one normally distributed, yielding one regression equation with three parameters, corresponding to an intercept and the two covariates. The expected response probability ranged from 20 to 75 percent yielding a minimum response category frequency of 10 events expected in a sample of 50 observations. A second series evaluated a binomial response in a regression with three binary and two normal covariates, with estimation of six parameters. The expected response probability ranged from 65 to 75 percent yielding a minimum response frequency of 12 events expected in 50 observations. The third series evaluated regressions on one binary and one normal covariate, but in this series the response was trinomial and therefore there were two regression equations, each with three parameters. The smaller of the two expected response probabilities ranged from 13 to 42 percent and the larger from 22 to 63 percent producing a minimum response frequency of 7 events expected in 50 observations. Within the bounds of 500–4000, we selected the number of datasets simulated for each combination of parameters such that the standard error of the mean bias of the uncorrected MLE was less than 1% of the largest slope parameter value.

Initially, we excluded datasets with complete, partial or quasi-separations (as defined in our previous paper), and when we detected a separation in the maximum likelihood estimation, we did not calculate bias reduced estimates by any method. The number of datasets with separations (i.e., with infinite estimates) was lowest in the binomial response models with two covariates, ranging between 0 and 12 percent in a sample size of 50. We therefore also conducted an additional set of simulations in a sample size of 25, allowing a maximum of 5000 datasets. In the trinomial models with a sample size of 50, 20–30 percent of datasets for three of the parameter combinations were initially excluded, but most had less than 10 percent exclusions. In the smaller sample sizes, the exclusion of datasets with infinite estimates tended to reduce the absolute value of the mean bias in the uncorrected MLEs. These summaries are interpreted therefore as conditional on the existence of all the uncorrected MLEs.

The complete set of simulations was repeated a second time, including those in which there were infinite MLEs, to evaluate the MPLEs over the entire sample space. We examined the distributions of the MPLEs in both separated and unseparated datasets using histograms and calculated the mean bias and MSE for the MPLEs.

We present tabled results for the simulations with correlated covariates in the smallest sample size evaluated in each of the three series of simulations. A complete set of tables that give mean bias and relative MSE results for all samples sizes and for both correlated and uncorrelated covariates is available upon request.

5.2. Results

As reported in previous studies, the MLEs have finite sample bias that is proportional to the true parameter value and increases in magnitude as the sample size decreases. Using an estimate of the asymptotic bias to correct the MLEs, however, appeared to correct beyond the true parameter value (Tables 4–6). In all cases studied, the BCMLEs had smaller MSE than the MLEs, but with severe overcorrection the MSE advantage of the BCMLEs begins to attenuate.

In the unseparated datasets, the mean bias of the MPLEs was smaller in magnitude than that of the BCMLEs, i.e., they did not overcorrect as much as the BCMLEs, particularly in the models with more parameters (Tables 5, 6). As the sample size increased, the mean bias of the MPLEs and the BCMLEs became more similar (data not shown). When the sample size was 200 they were indistinguishable, but both still had smaller bias and MSE than the MLEs. The MPLEs had MSE that was consistently less than that of the MLEs for both binary and normal covariates, primarily as a consequence of less variance in the MPLEs. The MSE of the MPLEs relative to the MLEs decreased as the sample size decreased and as the number of parameters being estimated increased, i.e., as the n-to-p ratio decreased. (Table 7).

In these simulations, it was usually one of the binary covariate parameter estimates that produced separation in the dataset, and apparently reasonable estimates could be obtained for the other parameters. The distribution of the MPLEs for the binary covariates in separated datasets includes larger values than those in unseparated datasets, whereas the distribution of the MPLEs of the normal covariates closely resembles that for the unseparated datasets (Fig. 1). This also explains the similarity of the mean bias of the MPLEs for the normal covariates in all datasets compared to unseparated datasets.

When all datasets were included in the summaries the mean bias of the MPLEs was usually closer to zero than that in the unseparated datasets, as a result of the parameters that had infinite-valued MLEs being given a large but finite MPLE value (Tables 4–6). Information from datasets in which there was a potentially strong association between the response and the binary covariate was thus rescued by the availability of the MPLEs in all datasets. In nearly all the simulation runs, for both binomial and trinomial responses, the mean bias of the MPLEs for all datasets was close to zero, and finite sample bias was effectively eliminated.

In some of our very sparse data simulations in which there was a substantial proportion of separated datasets (Tables 4 and 6), the mean bias of the MLEs for the binary covariates was observed to have a sign opposite to that expected (e.g., model 15 in Table 4, model 11 in Table 6). This occurred as an artefact of having to exclude separated datasets from the summary averages. For binary covariates with positive effects, this tended to exclude values from the upper tail of the distribution of estimates, producing a mean bias value that was shifted downward. The BCMLEs were also affected by this phenomenon, yielding shifted mean bias values particularly in very small samples. The minimal mean bias of the MPLEs averaged over both separated and unseparated datasets suggests that the apparent overcorrection of the MPLEs in the unseparated datasets (and a good part of that in the BCMLEs)

Table 4
Selected results for binomial logistic regression with two correlated[a] covariates ($n = 25$)

					Binary covariate (β_1)					Normal covariate (β_2)				
	# Simulations	Parameter values			Mean bias[c]			Relative MSE		Mean Bias[d]			Relative MSE	
Model[b]	(separated)	β_0	β_1	β_2	MLE[e]	BCMLE[f]	MPLE (∗)[g]	BCMLE	MPLE	MLE	BCMLE	MPLE (∗)	BCMLE	MPLE
13	5000 (919)	−1.4	0	1	−0.50	−0.26	−0.31 (−0.03)	0.42	0.44	0.39	−0.10	−0.10 (0.01)	0.32	0.33
14	5000 (435)	−1.4	1	0	0.02	−0.19	−0.15 (0.00)	0.58	0.64	0.02	0.01	0.01 (0.01)	0.49	0.56
15	5000 (949)	−1.4	1	1	−0.28	−0.37	−0.34 (−0.04)	0.55	0.60	0.37	−0.08	−0.08 (0.01)	0.37	0.35
16	5000 (173)	−1.4	1	−1	0.35	−0.05	0.02 (0.02)	0.45	0.52	−0.30	0.07	0.07 (0.00)	0.40	0.43
17	5000 (444)	−1.4	2	0	0.20	−0.24	−0.15 (0.01)	0.47	0.54	0.00	0.00	0.00 (0.01)	0.48	0.55
18	5000 (1045)	−1.4	2	1	−0.09	−0.50	−0.39 (−0.04)	0.57	0.61	0.40	−0.14	−0.14 (−0.02)	0.38	0.32
19	5000 (71)	0	0	1	−0.16	−0.04	−0.06 (−0.03)	0.53	0.59	0.33	−0.04	−0.04 (0.03)	0.39	0.41
20	5000 (122)	0	1	0	0.11	−0.09	−0.06 (−0.01)	0.59	0.65	0.01	0.00	0.00 (0.00)	0.54	0.60
21	5000 (410)	0	1	1	−0.04	−0.19	−0.15 (−0.01)	0.55	0.62	0.37	−0.08	−0.08 (0.01)	0.37	0.39
22	5000 (52)	0	1	−1	0.31	−0.05	0.02 (0.00)	0.51	0.52	−0.33	0.04	0.04 (−0.03)	0.40	0.38
23	5000 (1051)	0	2	0	0.01	−0.44	−0.34 (−0.01)	0.58	0.61	−0.02	−0.02	−0.02 (−0.02)	0.49	0.55
24	5000 (1705)	0	2	1	−0.32	−0.67	−0.58 (−0.09)	0.69	0.69	0.41	−0.14	−0.14 (−0.02)	0.28	0.31

[a] The correlation between the binary and the normal covariate is 0.6.
[b] Models 1–12 (not shown) correspond to models with the same parameters, but with uncorrelated covariates.
[c] The standard errors of the mean bias of the MLE for the binary covariate range from 0.021 to 0.029; they are smaller for BCMLE and MPLE.
[d] The standard errors of the mean bias of the MLE for the normal covariate range from 0.016 to 0.026; they are smaller for BCMLE and MPLE.
[e] MLE: maximum likelihood estimate in unseparated datasets.
[f] BCMLE: asymptotic bias corrected estimate in unseparated datasets.
[g] MPLE: modified score function estimates in unseparated datasets; mean bias in all datasets, including separated, is given in brackets.

Table 5
Selected results for binomial logistic regression with five covariates—three binary and two normal covariates ($n = 50$). In Models 1–3, covariates are uncorrelated; in models 4–6, covariates are correlated[a]

			Mean bias[c]			Relative MSE					Mean bias			Relative MSE	
Model[b]	# Simulations (separated)	ß	MLE[d]	BCMLE[e]	MPLE (∗)[f]	BCMLE	MPLE	Model	# Simulations (separated)	ß	MLE	BCMLE	MPLE (∗)	BCMLE	MPLE
1	2769 (17)	1	0.30	−0.04	0.02 (0.03)	0.42	0.49	4	4000 (214)	1	0.29	−0.13	−0.02 (0.01)	0.31	0.40
		2	0.59	−0.09	0.05 (0.05)	0.32	0.36			2	0.68	−0.29	−0.05 (0.03)	0.29	0.33
		1	0.29	−0.05	0.01 (0.01)	0.44	0.49			1	0.30	−0.11	−0.01 (−0.01)	0.31	0.40
		1	0.33	−0.04	0.03 (0.03)	0.33	0.37			1	0.39	−0.14	−0.01 (0.02)	0.30	0.36
		1	0.33	−0.05	0.02 (0.02)	0.29	0.36			1	0.40	−0.12	0.00 (0.03)	0.30	0.37
2	2908 (16)	1	0.30	−0.05	0.01 (0.02)	0.39	0.47	5	3664 (309)	1	0.23	−0.05	0.01 (0.02)	0.63	0.49
		2	0.61	−0.09	0.03 (0.06)	0.28	0.36			2	0.44	−0.22	−0.08 (0.04)	0.63	0.41
		1	0.28	−0.06	0.00 (0.00)	0.43	0.50			1	0.18	−0.07	−0.02 (−0.04)	0.41	0.50
		−1	−0.34	−0.04	−0.02 (−0.03)	0.28	0.35			−1	−0.33	0.06	−0.01 (−0.02)	0.63	0.44
		1	0.34	−0.04	0.02 (0.03)	0.28	0.36			1	0.32	−0.07	0.00 (0.01)	0.94	0.44
3	3660 (93)	1	0.27	−0.13	−0.05 (−0.03)	0.36	0.44	6	4000 (485)	1	0.33	−0.16	−0.03 (0.01)	0.30	0.36
		2	0.62	−0.21	−0.04 (0.01)	0.31	0.32			2	0.67	−0.51	−0.14 (0.01)	0.40	0.29
		2	0.62	−0.22	−0.05 (0.01)	0.31	0.33			2	0.71	−0.45	−0.10 (0.04)	0.41	0.29
		1	0.33	−0.10	−0.02 (0.00)	0.26	0.32			1	0.43	−0.23	−0.04 (0.02)	0.41	0.31
		1	0.35	−0.10	0.00 (0.01)	0.33	0.32			1	0.41	−0.22	−0.04 (0.02)	0.36	0.33

[a]The pairwise correlations are 0.6 for the binary covariates; 0.8 for the normal covariates; and 0.0 between the binary and the normal covariates.
[b]The intercept is −1.4 in all models.
[c]The standard errors of the mean bias of the MLE for the binary covariates range from 0.024 to 0.051; those for the mean bias of the MLE for the normal covariates range from 0.014 to 0.025; they are smaller for BCMLE and MPLE.
[d]MLE: maximum likelihood estimate in unseparated datasets.
[e]BCMLE: asymptotic bias corrected estimate in unseparated datasets.
[f]MPLE: modified score function estimates in unseparated datasets; mean bias in all datasets, including separated, is given in brackets.

Table 6
Selected results for trinomial logistic regression with two correlated[a] covariates ($n = 50$)

		Parameter values			Binary (β_{11}, β_{21})					Normal (β_{12}, β_{22})				
					Mean bias[c]			Relative MSE		Mean bias[d]			Relative MSE	
	# Simulations	β_{10}	β_{11}	β_{12}										
Model[b]	(separated)	β_{20}	β_{21}	β_{22}	MLE[e]	BCMLE[f]	MPLE (∗)[g]	BCMLE	MPLE	MLE	BCMLE	MPLE (∗)	BCMLE	MPLE
8	2564 (214)	−1.4	0	1	−0.21	−0.08	−0.09 (0.00)	0.56	0.62	0.22	−0.02	0.00 (0.01)	0.50	0.56
		−1.4	2	0	0.17	−0.05	−0.02 (0.02)	0.72	0.75	0.01	0.00	0.00 (0.00)	0.75	0.77
9	4000 (324)	−1.4	1	1	−0.06	−0.18	−0.15 (−0.02)	0.72	0.75	0.19	−0.01	0.01 (0.02)	0.61	0.64
		−1.4	1	−1	0.14	0.00	0.01 (0.00)	0.70	0.73	−0.16	0.02	0.00 (0.00)	0.63	0.66
10	4000 (833)	−1.4	2	1	−0.13	−0.46	−0.39 (−0.05)	0.77	0.79	0.25	0.00	−0.02 (0.02)	0.54	0.58
		0	1	−1	0.13	−0.07	−0.04 (0.00)	0.70	0.74	−0.13	0.03	0.02 (0.01)	0.68	0.70
11	3681 (1005)	−1.4	2	1	−0.25	−0.61	−0.53 (−0.07)	0.75	0.77	0.20	−0.01	0.01 (0.00)	0.57	0.60
		0	2	1	−0.09	−0.36	−0.31 (0.05)	0.80	0.82	0.18	−0.02	0.00 (−0.01)	0.57	0.60
12	3424 (281)	0	0	1	−0.23	−0.20	−0.19 (−0.05)	0.68	0.70	0.19	0.00	0.02 (0.02)	0.63	0.66
		0	2	0	0.07	−0.20	−0.16 (−0.02)	0.73	0.75	0.00	−0.01	−0.01 (0.01)	0.70	0.73
13	3557 (308)	0	1	0	0.01	−0.18	−0.15 (0.00)	0.71	0.75	−0.02	−0.01	−0.01 (−0.02)	0.72	0.74
		0	2	0	0.07	−0.21	−0.17 (−0.02)	0.72	0.75	0.00	−0.01	−0.01 (−0.01)	0.72	0.74
14	4000 (1203)	0	2	1	−0.32	−0.62	−0.56 (−0.11)	0.94	0.92	0.18	−0.02	−0.02 (0.01)	0.62	0.64
		0	2	1	−0.30	−0.60	−0.53 (−0.11)	0.93	0.91	0.18	−0.02	0.00 (0.00)	0.61	0.64

[a]The correlation between the binary and the normal covariate is 0.6.
[b]Models 1–7 (not shown) correspond to models with the same parameter values as models 8–14 but uncorrelated covariates.
[c]The standard errors of the mean bias of the MLE for the binary covariate range from 0.019 to 0.026; they are smaller for BCMLE and MPLE.
[d]The standard errors of the mean bias of the MLE for the normal covariate range from 0.012 to 0.020; they are smaller for BCMLE and MPLE.
[e]MLE: maximum likelihood estimate in unseparated datasets.
[f]BCMLE: asymptotic bias corrected estimate in unseparated datasets.
[g]MPLE: modified score function estimates in unseparated datasets; mean bias in all datasets, including separated, is given in brackets.

Table 7
Range of relative mean squared error (RMSE) of the MPLE to the uncorrected MLE (percent)

Sample size	Binomial regression		Trinomial regression
	Two covariates	Five covariates	Two covariates
200	89–96	81–92	83–95
100	78–93	61–79	75–89
75	76–90	44–71	71–87
50	62–84	27–52	55–92
25	28–69	—	—

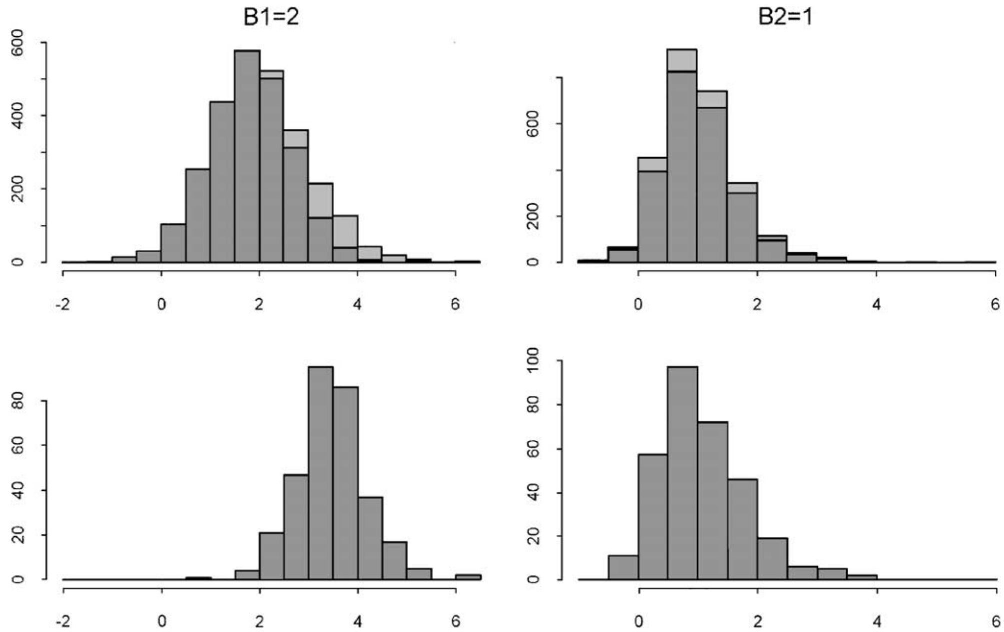

Fig. 1. Distribution of MPLEs for binomial logistic regression with one binary (x_1) and one normal covariate (x_2) in 2719 simulations of a dataset of 50 observations. The model is logit $(\theta) = \alpha + \beta_1 x_1 + \beta_2 x_2$ with $\alpha = 0, \beta_1 = 2, \beta_2 = 1$. The upper panels include estimates from all datasets; datasets with at least one infinite MLE are denoted by the light grey histogram bars. The lower panels include only the 315 datasets with separations.

may be the result of averaging over the incomplete sample space in which the MLEs are finite.

6. Discussion

The applications and the Monte Carlo study demonstrate several advantages for the modified score function estimator. In the applications, it is apparent that the shrinkage effect of the modified score function operated more strongly when it was needed, i.e., when there was a large association and the ratio of observations to parameters was small, but was minimal for parameter estimates that did not require bias reduction. It was effectively applied in both very small samples and in large samples in which sparse data made estimation of some of the parameters difficult. It was possible to model multiple quantitative covariates without having to arbitrarily group values into a smaller number of levels. In the simulation studies, the bias reduction was equally effective for binomial and trinomial responses and for regressions with general covariate types, including continuous covariates.

In previous studies of methods for bias reduction, our comparisons were limited to datasets with no infinite estimates and we observed that as the sample size decreased, exclusion of such datasets produced progressive truncation of one tail of the distribution of the MLEs of large parameters. As a result, in some simulations the mean bias initially increased but eventually decreased in the smallest sample size. In this study, we were able to observe more directly the effects on the bias reduced estimates of limiting consideration to the unseparated datasets, because the MPLEs were calculated in both unseparated and separated datasets. Thus, some of the apparent overcorrection observed in evaluations of the bias reduction methods that require the existence of the MLEs was likely due to consideration of an incomplete sample space.

In the smallest sample simulations, the MPLEs were less biased and more efficient than the unconditional MLEs and the bias-corrected MLEs, and could be obtained in all samples, even when there were infinite MLEs. In moderate samples, the MPLEs were less biased and more efficient than the unconditional MLEs, and behaved similarly to the bias-corrected MLEs. As the sample size increases, the modified estimates become equivalent to the usual MLEs, so routine application of modified estimation appears to bear only the cost of implementation and additional computation. These results encourage us to develop and evaluate computationally efficient algorithms for hypothesis testing and confidence interval methods based on the modified score function and corresponding penalized log likelihood.

A referee has noted that the score modification could be generalized by using priors stronger than Jeffreys', further reducing MSE at the cost of introducing negative bias on the log-odds-ratio scale. Greenland (2000) demonstrates this for the conditional logistic likelihood. The penalized log likelihood, l^*, that yields the modified score, U^*, is the log posterior density obtained when the log prior density, $\frac{1}{2}\log|A|$, is the anti-derivative of the score penalty. More generally, then, a log prior density of $c\log|A|$ with $c > \frac{1}{2}$ yields a larger score penalty; for example, in tabular data, $c = 1$ corresponds to the Laplace estimate in which 1 instead of $\frac{1}{2}$ is added per cell. Hauck et al. (1982) examined alternative values of c in the context of a series of 2×2 tables, and further work for regression models is warranted. References Hirji et al., 1989; SAS/STAT, 1996

Acknowledgements

During this study, S.B. Bull was a National Health Research Scholar of the National Health Research and Development Program of Health and Welfare Canada.

This research was supported by the Natural Sciences and Engineering Research Council of Canada. Thanks to J.P. Lewinger for assistance in preparing the applications and for helpful discussions.

References

Albert, A., Anderson, J.A., 1984. On the existence of maximum likelihood estimates in logistic regression models. Biometrika 71, 1–10.

Albert, A., Harris, E.K., 1984. Multivariate Interpretation of Clinical Laboratory Data. Dekker, New York.

Alho, J.M., 1992. On the computation of likelihood and score test based confidence intervals in generalized linear models. Statist. Med. 11, 923–990.

Anderson, J.A., Richardson, S.C., 1979. Logistic discrimination and bias correction in maximum likelihood estimation. Technometrics 21, 71–78.

Aptech Systems Incorporated, 1990. The GAUSS System, Version 2.0, Kent, Washington.

Blajchman, M.A., Bull, S.B., Feinman, S.V., 1995. For the Canadian post-transfusion hepatitis prevention study group. Post-transfusion hepatitis: impact of the non-A non-B hepatitis surrogate tests. Lancet 345, 21–25.

Brown, B.W., 1980. Prediction analyses for binary data. In: Miller, R.G., Efron, B., Brown, B.W., Moses, L.E. (Eds.), Biostatistics Casebook. Wiley, New York, pp. 3–18.

Bull, S.B., Hauck, W.W., Greenwood, C.M.T., 1994. Two-step jackknife bias reduction for logistic regression MLEs. Commun. Statist.—Simulation Comput. 23, 59–88.

Bull, S.B., Greenwood, C.M.T., Hauck, W.W., 1997. Jackknife bias reduction for polychotomous logistic regression. Statist. Med. 16, 545–560.

Copas, J.B., 1988. Binary regression models for contaminated data, with discussion. J. Royal Statist. Soc. Ser. B 50, 225–265.

Cordeiro, G.M., McCullagh, P., 1991. Bias correction in generalized linear models. J. Royal Statist. Soc. Ser. B 53, 629–643.

Cox, D.R., Snell, E.J., 1968. A general definition of residuals. J. Royal Statist. Soc. Ser. B 30, 248–275.

Cox, D.R., Snell, E.J., 1989. The Analysis of Binary Data., 2nd Edition. Chapman and Hall, New York.

Cytel Software Corporation, 1992. LogXact-Turbo: A Software Package for Exact and Asymptotic Logistic Regression, Version 1.0, Cambridge, Massachusetts.

Duffy, D.E., Santner, T.J., 1989. On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models. Commun. Statist.—Theory Meth. 18, 959–980.

Farewell, V.T., 1978. Jackknife estimation with structured data. Biometrika 65, 444–447.

Firth, D., 1992a. Generalized linear models and Jeffreys priors: an iterative weighted least-squares approach. In: Dodge, Y., Whittaker, J. (Eds.), Computational Statistics, Vol. 1. Physica-Verlag, Heidelberg, pp. 553–557.

Firth, D., 1992b. Bias reduction, the Jeffreys prior and GLIM. In: Fahrmeir, L., Francis, B., Gilchrist, R., Tutz, G. (Eds.), Advances in GLIM and Statistical Modelling. Springer, New York, pp. 91–100.

Firth, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika 80, 27–38.

Greenland, S., 1995. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. Epidemiology 6, 450–454.

Greenland, S., 2000. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. Biostatistics 1, 113–122.

Greenland, S., Schwartzbaum, J.A., Finkle, W.D., 2000. Problems due to small samples and sparse data in conditional logistic regression analysis. Amer. J. Epidemiol. 151, 531–539.

Haldane, J.B.S., 1956. The estimation and significance of the logarithm of a ratio of frequencies. Ann. Human Genetics 20, 309–311.

Hauck, W.W., 1984. A comparative study of conditional maximum likelihood estimation of a common odds ratio. Biometrics 40, 1117–1123.

Hauck, W.W., Donner, A., 1977. Wald's test as applied to hypothesis testing in logit analysis. J. Amer. Statist. Assoc. 72, 851–853.

Hauck, W.W., Anderson, S., Leahy, F.J., 1982. Finite-sample properties of some old and some new estimators of a common odds ratio from multiple 2×2 tables. J. Amer. Statist. Assoc. 77, 145–152.

Hirji, K.F., 1992. Computing exact distributions for polytomous response data. J. Amer. Statist. Assoc. 87, 487–492.

Ibrahim, J.G., Laud, P.W., 1991. On Bayesian analysis of generalized linear models using Jeffreys's prior. J. Amer. Statist. Assoc. 86, 981–986.

Jennings, D.E., 1986. Judging inference adequacy in logistic regression. J. Amer. Statist. Assoc. 81, 471–476.

Lesaffre, E., Albert, A., 1989. Partial separation in logistic discrimination. J. Royal Statist. Soc. Ser. B 51, 109–116.

Mehta, C.R., Patel, N.R., 1995. Exact logistic regression: theory and examples. Statist. Med. 14, 2143–2160.

Pettitt, A.N., Kelly, J.M., Gao, J.T., 1998. Bias correction for censored data with exponential lifetimes. Statist. Sinica 8, 941–963.

Rubin, D.B., Schenker, N., 1987. Logit-based interval estimation for binomial data using Jeffreys prior. In: Clogg, C.C. (Ed.), Sociological Methodology 1987. American Sociological Association, Washington, DC, pp. 131–144.

Santner, T.J., Duffy, D.E., 1986. A note on A. Albert and J.A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. Biometrika 73, 755–758.

Santner, T.J., Duffy, D.E., 1989. The Statistical Analysis of Discrete Data. Springer, New York.

SAS/STAT, 1996. Proc Logistic, release 6.12, Cary, NC: SAS Institute.

Schaefer, R.L., 1983. Bias correction in maximum likelihood logistic regression. Statist. Med. 2, 71–78.

Venzon, D.J., Moolgavkar, S.H., 1988. A method for computing profile-likelihood-based confidence intervals. Appl. Statist. 37, 87–94.

Zelen, M., 1991. Multinomial response models. Comput. Statist. Data Anal. 12, 249–254.