Research Article

Statistics in Medicine

Received 14 November 2008, Accepted 8 September 2010 Published online 6 November 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4105

Bias reduction in conditional logistic regression

Jenny X. Sun,[a] Samiran Sinha,[a*†] Suojin Wang[a] and Tapabrata Maiti[b]

We employ a general bias preventive approach developed by Firth (*Biometrika* 1993; 80:27–38) to reduce the bias of an estimator of the log-odds ratio parameter in a matched case–control study by solving a modified score equation. We also propose a method to calculate the standard error of the resultant estimator. A closed-form expression for the estimator of the log-odds ratio parameter is derived in the case of a dichotomous exposure variable. Finite sample properties of the estimator are investigated via a simulation study. Finally, we apply the method to analyze a matched case–control data from a low birthweight study.

Keywords: bias; conditional logistic regression; jackknife method; matched case–control study; modified score function

1. Introduction

Conditional likelihood is widely used for the estimation of log-odds ratio from matched case–control studies where a case (read diseased) subject is matched with a number of controls (read non-diseased) based on some matching or confounding factors. Matched designs are commonly used in the situation where both the disease probability and the exposure of interest depend on a common set of variables which are used as matching variables. In the conditional likelihood analysis of a matched design, we remove the stratum-specific nuisance parameters by conditioning on their sufficient statistics, and obtain a consistent estimate of the log-odds ratio parameter by maximizing the conditional likelihood. In a matched case–control study, when the sample size is not large, the maximum conditional likelihood (MCL) estimator obtained by maximizing the conditional likelihood is generally biased. Thus it is important to develop a method that can produce an estimate of the parameters of interest with little or less bias. For instance, the dataset that motivated our research is from a low birthweight study, consisting of only 29 strata, with each stratum having only one case and three controls.

There are two main types of approaches that deal with the bias of a point estimator: a bias *corrective* method and a bias *preventive* method. In the corrective method the resultant estimator called a bias-corrected estimator depends on the existence of the maximum likelihood estimator and is obtained by subtracting the first-order term in the asymptotic expansion of the bias from the maximum likelihood estimator. Within the corrective methods, Jewell [1] considered a computationally intensive jackknife (JNF) method for correcting the bias in the odds ratio estimator for a categorical exposure variable and compared the performance of his proposed method with some other bias correction approaches. He exclusively focused on a single exposure variable with finite categories. Several other bias correction techniques for the odds-ratio estimator for discrete exposure variables were proposed by Bishop *et al.* [2]. Greenland [3] adopted the general bias correction approach of Cordeiro and McCullagh [4] for correcting bias in the conditional logistic regression setting. One main drawback of the corrective approach is that if the maximum likelihood estimate has an infinite component, which is not uncommon for small sample sizes, then the bias corrected estimator is undefined. The JNF method of bias correction is even more likely to encounter the same problem. To avoid this difficulty, Firth [5] proposed the bias preventive method. This approach eliminates the first-order of bias $O(n^{-1})$ by solving a modified score equation in the general context of regular parametric families of distributions. Therefore the resultant estimator does not depend on the finiteness of the classical maximum likelihood estimator. Firth's idea has been used in the unconditional logistic regression [6–8] and Cox's partial likelihood setting to handle the problem of monotone likelihood [9]. In this paper we will use this bias preventive procedure in the context of conditional logistic regression models and derive the

[a]*Department of Statistics, Texas A&M University, College Station, TX 77843, U.S.A.*
[b]*Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, U.S.A.*
[*]*Correspondence to: Samiran Sinha, Department of Statistics, Texas A&M University, College Station, TX 77843, U.S.A.*
[†]*E-mail: sinha@stat.tamu.edu*

modified score function to estimate the model parameters. We refer to the method as a modified score function (MDS) approach.

The novel features of this work are highlighted as follows. First, to the best of our knowledge this is the first bias preventive approach in the MCL estimator in matched case–control studies. Second, compared to the JNF or bootstrap-based procedures the MDS method takes much less computational time and effort. Third, the MDS estimates are usually finite even when MCL estimates are infinite. Fourth, we provide a versatile formula for standard error calculation for the MDS estimator. Furthermore, for small sample sizes our numerical studies indicate that the MDS estimator not only reduces a substantial amount of bias of the estimate but also usually has a smaller variance than that of the MCL estimator.

2. Model and assumptions

Suppose we have an $1{:}M_i(\geqslant 1)$ matched case–control data with n strata. Let Y_{ij} take on value 1 or 0 accordingly as the jth subject in the ith matched set is a case or control, respectively. Let $\boldsymbol{X}_{ij}=(X_{ij1},\ldots,X_{ijp})^\top$ be a $p\times 1$ vector of covariates. Also, let $\boldsymbol{S}_i$ be the covariates which are used for matching purposes in the ith stratum. Let the disease risk model for the ith stratum be

$$pr(Y_{ij}=1|\boldsymbol{S}_i,\boldsymbol{X}_{ij})=H(\alpha_i(\boldsymbol{S}_i)+\boldsymbol{X}_{ij}^\top\boldsymbol{\beta})\quad \text{for } j=1,\ldots,M_i+1,\ i=1,\ldots,n \tag{1}$$

with $H(z)=\{1+\exp(-z)\}^{-1}$. Note that α_i is the stratum-specific parameter which is a function of $\boldsymbol{S}_i$ and $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)^\top$ is the vector of log-odds ratio parameters for the covariate $\boldsymbol{X}_{ij}$. The MDS method is applicable for a matched case–control study with varying number of controls in each stratum as long as M_is are bounded as $n\to\infty$. In order to estimate $\boldsymbol{\beta}$ in equation (1) one generally adopts the conditional logistic regression [10] where the estimates are obtained by maximizing

$$L_{\text{CLR}}(\boldsymbol{\beta})=\prod_{i=1}^{n}\sum_{j=1}^{M_i+1}p_{ij}Y_{ij}, \tag{2}$$

where $p_{ij}=\exp(\boldsymbol{X}_{ij}^\top\boldsymbol{\beta})/\sum_{k=1}^{M_i+1}\exp(\boldsymbol{X}_{ik}^\top\boldsymbol{\beta})$ representing the conditional probability that the jth subject is a case given that there is one case in the ith stratum. To obtain L_{CLR} we condition on $\sum_{j=1}^{M_i+1}Y_{ij}=1$ which is a sufficient statistic for $\alpha_i(\boldsymbol{S}_i)$. The MCL estimator for $\boldsymbol{\beta}$ is obtained by solving the score equation $U(\boldsymbol{\beta})=\partial\log(L_{\text{CLR}})/\partial\boldsymbol{\beta}=\sum_{i=1}^{n}\sum_{j=1}^{M_i+1}(Y_{ij}-p_{ij})\boldsymbol{X}_{ij}=0$, and Fisher's information matrix is $I(\boldsymbol{\beta})=\sum_{i=1}^{n}\{\sum_{j=1}^{M_i+1}\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^\top p_{ij}-(\sum_{j=1}^{M_i+1}\boldsymbol{X}_{ij}p_{ij})(\sum_{j=1}^{M_i+1}\boldsymbol{X}_{ij}p_{ij})^\top\}=\sum_{i=1}^{n}\sum_{j=1}^{M_i+1}(\boldsymbol{X}_{ij}-\overline{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{ij}-\overline{\boldsymbol{X}}_{i\cdot})^\top p_{ij}$, where $\overline{\boldsymbol{X}}_{i\cdot}=\sum_{j=1}^{M_i+1}\boldsymbol{X}_{ij}p_{ij}$.

3. Method of bias reduction

Since conditional logistic regression is a member of the full exponential family of distributions, we apply Firth's result related to the exponential family that penalizes the likelihood by Jeffreys invariant prior to yield estimators free of first-order bias. Let the modified conditional score equation be

$$U^{\text{mod}}(\boldsymbol{\beta})=(U_{(1)}^{\text{mod}}(\boldsymbol{\beta}),\ldots,U_{(p)}^{\text{mod}}(\boldsymbol{\beta}))^\top=\boldsymbol{0}, \tag{3}$$

where the rth component of the modified conditional score function

$$\begin{aligned}U_{(r)}^{\text{mod}}(\boldsymbol{\beta})&=U_{(r)}(\boldsymbol{\beta})+\frac{1}{2}\frac{\partial}{\partial\beta_r}\{\log|I(\boldsymbol{\beta})|\}=U_{(r)}(\boldsymbol{\beta})+\frac{1}{2}tr\left\{I^{-1}(\boldsymbol{\beta})\frac{\partial I(\boldsymbol{\beta})}{\partial\beta_r}\right\}\\&=\sum_{i=1}^{n}\sum_{j=1}^{M_i+1}(Y_{ij}-p_{ij})X_{ijr}+\frac{1}{2}tr\left[\left\{\sum_{i=1}^{n}\sum_{j=1}^{M_i+1}(\boldsymbol{X}_{ij}-\overline{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{ij}-\overline{\boldsymbol{X}}_{i\cdot})^\top p_{ij}\right\}^{-1}\right.\\&\quad\left.\times\left\{\sum_{i=1}^{n}\sum_{j=1}^{M_i+1}(\boldsymbol{X}_{ij}-2\overline{\boldsymbol{X}}_{i\cdot})\boldsymbol{X}_{ij}^\top(X_{ijr}-\overline{X}_{i\cdot r})p_{ij}\right\}\right],\end{aligned}$$

$U(\boldsymbol{\beta})=(U_{(1)}(\boldsymbol{\beta}),\ldots,U_{(p)}(\boldsymbol{\beta}))^\top$, and $\overline{X}_{i\cdot r}=\sum_{j=1}^{M_i+1}X_{ijr}p_{ij}$. Following the argument in Firth [5], it is seen that the adjustment term after the standard conditional score function above effectively eliminates the first-order bias of the conditional maximum likelihood estimator. Note that we can write the rth component of the MDS $U^{\text{mod}}_{(r)}(\boldsymbol{\beta})=\sum_{i=1}^{n} U^{\text{mod}}_{(r)i}(\boldsymbol{\beta})$, where $U^{\text{mod}}_{(r)i}(\boldsymbol{\beta})=\sum_{j=1}^{M_i+1}(Y_{ij}-p_{ij})\boldsymbol{X}_{ij}+tr\{I^{-1}(\boldsymbol{\beta})(\partial I(\boldsymbol{\beta})/\partial\beta_r)\}/(2n)$. In addition, define $U^{\text{mod}}_{i}(\boldsymbol{\beta})=(U^{\text{mod}}_{(1)i}(\boldsymbol{\beta}),\ldots,U^{\text{mod}}_{(p)i}(\boldsymbol{\beta}))^{\top}$. Let $\widehat{\boldsymbol{\beta}}_{\text{R}}$ be the MDS estimator obtained by solving equation (3).

Although according to the development in Firth [5] the modified score estimator has the same asymptotic variance–covariance matrix as the MCL estimator, in the simulation study we found that for small to moderate sample sizes the sandwich type estimator

$$\widehat{\text{var}}(\widehat{\boldsymbol{\beta}}_{\text{R}})=\left[\left(\frac{\partial U^{\text{mod}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\top}}\right)^{-1}\left(\sum_{i=1}^{n}U_i^{\text{mod}}(\boldsymbol{\beta})U_i^{\text{mod}^{\top}}(\boldsymbol{\beta})\right)\left(\frac{\partial U^{\text{mod}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\top}}\right)^{-\text{T}}\right]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{\text{R}}} \tag{4}$$

generally yields a more accurate estimate of the true standard error (TSD) (approximated by the empirical standard error in the tables) than that obtained by inverting Fisher's information matrix which ignores the correction term of the MDS. One explanation for better accuracy of formula (4) is that it takes into account the correction term of the MDS which is of order $O(1)$ and may not be negligible for small n. The JNF method of asymptotic variance estimation is not only time consuming but may also produce inaccurate variance estimates. Thus, in this respect the MDS method not only provides an approach to handle bias but is also accompanied by an explicit and useful formula for the estimation of the asymptotic variance. The expression of $\partial U^{\text{mod}}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^{\top}=\{\partial U^{\text{mod}}_{(r)}(\boldsymbol{\beta})/\partial\beta_k\}$ is given in the Appendix.

Example 1: *The case of a single covariate*

For a single covariate the first-order MDS estimator of the log-odds ratio parameter is obtained by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{M_i+1}(Y_{ij}-p_{ij})X_{ij}+\frac{\sum_{i=1}^{n}\sum_{j=1}^{M_i+1}X_{ij}(X_{ij}-\overline{X}_{i\cdot})(X_{ij}-2\overline{X}_{i\cdot})p_{ij}}{2\sum_{i=1}^{n}\sum_{j=1}^{M_i+1}(X_{ij}-\overline{X}_{i\cdot})^2p_{ij}}=0. \tag{5}$$

The first term on the left-hand side above is the score function derived from the conditional likelihood (2) and the second term is the correction term. Consider an $1:M_i$ matched study with $j=1$ representing the cases and otherwise controls. When $X_{i1}>0$ and $X_{ij}<0$ for $j=2,\ldots,M_i+1$, the score equation derived from the conditional likelihood is $\sum_{i=1}^{n}(1-p_{i1})X_{i1}=\sum_{i=1}^{n}\sum_{j=2}^{M_i+1}p_{ij}X_{ij}$. It is clearly seen that the left-hand side of the equation is positive whereas the right-hand side is negative. Thus there is no finite solution for β. In contrast, when we use equation (5) we do not experience such a problem in our numerical computations.

Example 2: *Matched pair design with a dichotomous exposure*

Now we consider the matched pair design, i.e. $M_i=1$ for $i=1,2\ldots,n$, and with the exposure X taking on 0 or 1. Let u and v be the number of discordant matched pairs where the case is exposed and control is unexposed, and the case is unexposed and the control is exposed, respectively. The MCL estimator of the log-odds ratio is $\widehat{\beta}=\log(u/v)$ with asymptotic variance $\text{var}(\widehat{\beta})=1/\{(u+v)H(\beta)[1-H(\beta)]\}$. A consistent estimator of $\text{var}(\widehat{\beta})$ is $\widehat{\text{var}}(\widehat{\beta})=(u+v)/(uv)$. In this case Greenland's bias corrected estimator of β is given by $\widehat{\beta}_{\text{G}}=\log(u/v)-(u-v)^2/\{2uv(u+v)\}$. No explicit formula for its asymptotic variance was given although it could be approximated by the delta method as indicated in Greenland [3].

In the MDS method we estimate β by solving $U^{\text{mod}}(\beta)=U(\beta)+(1/2)I^{-1}(\beta)\partial I(\beta)/\partial\beta=0$, where $U(\beta)=u-(u+v)H(\beta)$, $I(\beta)=(u+v)H(\beta)\{1-H(\beta)\}$, and $\partial I(\beta)/\partial\beta=(u+v)H(\beta)\{1-H(\beta)\}\{1-2H(\beta)\}$. The MDS estimator becomes

$$\widehat{\beta}_{\text{R}}=\log\left(\frac{2u+1}{2v+1}\right).$$

Note that for a small sample size the MCL estimate could be infinite. Clearly, it is more likely for the JNF estimator to have this difficulty. However, as we see $\widehat{\beta}_{\text{R}}$ will not be infinite even for $v=0$.

Furthermore, $\partial U^{\text{mod}}(\beta)/\partial\beta=-(1+u+v)H(\beta)\{1-H(\beta)\}$ and $\sum_{i=1}^{n}\{U_i^{\text{mod}}(\beta)\}^2=1/(4n)\{1-2H(\beta)\}^2+u\{1-H(\beta)\}^2+vH^2(\beta)+1/n(u-v)H(\beta)\{1-H(\beta)\}$. Here n is the total number of all matched pairs, thus $n\geqslant u+v$. Then a consistent estimator of $\text{var}(\widehat{\beta}_{\text{R}})$ is

$$\widehat{\text{var}}(\widehat{\beta}_{\text{R}})=4\left\{\frac{u}{(1+2u)^2}+\frac{v}{(1+2v)^2}-\frac{(u-v)^2}{n(1+2u)^2(1+2v)^2}\right\}. \tag{6}$$

It can be directly shown that the above variance formula is non-negative. Since $\widehat{\beta}_{\text{R}}$ implicitly involves n, $\widehat{\text{var}}(\widehat{\beta}_{\text{R}})$ is a function of n. In addition, if one carries out a binomial experiment with $u+v$ trials, and u successes are observed, then $\widehat{\beta}_{\text{R}}$ becomes identical to the modified empirical estimator of the logit of the success probability of the binomial experiment given in Cox and Snell [11, Section 2.1.9]. This fact was also recognized by Firth [5] in the context of unconditional logistic regression. It was also implied for conditional logistic regression for 1:1 matched case–control study as the latter is equivalent to the unconditional logistic regression with appropriately defined covariates. Let $u/(u+v)\to\rho$ as $n\to\infty$ then one can obtain that $\widehat{\text{var}}(\widehat{\beta}_{\text{R}})/\{(u+v)\rho(1-\rho)\}^{-1}\to 1$. That implies $\widehat{\text{var}}(\widehat{\beta}_{\text{R}})=\{(u+v)\rho(1-\rho)\}^{-1}+o_p(1)$. Therefore, although our asymptotic variance estimator $\widehat{\text{var}}(\widehat{\beta}_{\text{R}})$ is not the same as the variance estimator given in Cox and Snell [11], they are asymptotically first-order equivalent.

Following a comment by a reviewer we discuss the scenario when $u=0$, $u+v=n$ (=5, say for example). Thus, the data do not have any concordant pairs. In this scenario $\widehat{\beta}_{\text{R}}=\log(1/11)$ and the estimated variance based on formula (6) will be zero. We explain this situation as follows. First, our variance formula is asymptotically correct. Second, for large n, the probability of getting $u=0$ is negligible. Thus, in this asymptotically unlikely scenario, we do not recommend to use formula (6).

4. A simulation study

In order to judge the performance of the methods we conducted a simulation study with the following three scenarios. We first generated cohort data consisting of three variables S, X, and Y according to the simulation scenarios 1 and 2, and with four variables S, X, Z, and Y according to simulation scenario 3.

1. 1:1 matched case–control data with $S\sim\text{Normal}(0.53, 0.24^2)$, $X\sim\text{Bernoulli}(p_x)$, $p_x=H(-2.0+S)$, and marginally $p_x\approx 0.2$, and $Y\sim\text{Bernoulli}(p_y)$, where $p_y=H(\beta_0+1.1S+\beta X)$; with $\beta_0=-2.9$ and $\beta=0.5, 1$.
2. 1:2 matched case–control data with $S\sim\text{Normal}(0.53, 0.24^2)$, $X\sim\text{Gamma}(S^2+0.5, 2.5)$, and $Y\sim\text{Bernoulli}(p_y)$, where $p_y=H(\beta_0+1.1S+\beta X)$; with $\beta_0=-3.0$ for $\beta=0.5$, 1, $\beta_0=-3.4$ for $\beta=1.5$, and $\beta_0=-3.7$ for $\beta=2.0$.
3. 1:1 matched case–control data with $S\sim\text{Normal}(0.53, 0.24^2)$, $Z\sim\text{Bernoulli}(p_z)$, $p_z=H(-1.7+1.3S)$, $X\sim\text{Bernoulli}(p_x)$, $p_x=H(-2.4+1.4S+1.0Z)$, and $Y\sim\text{Bernoulli}(p_y)$, where $p_y=H(\beta_0+1.1S+\beta_1 Z+\beta_2 X)$; with $\beta_0=-3.0$ for $(\beta_1,\beta_2)=(0.5, 0.5)$ and $\beta_0=-3.3$ for $(\beta_1,\beta_2)=(1,1)$.

The distribution and log-odds ratio parameter for S in p_y were chosen by mimicking a dataset that we analyzed. We chose a value of β_0 so that the overall marginal disease prevalence is around 10 per cent. The association between X and S is small to moderate, such as $\text{corr}(X,S)=0.1$, 0.28, and 0.15 for the three scenarios, respectively.

From the cohort data we created $1:M$ (here we chose $M_i=M$) matched case–control data with n strata using S as the matching variable. For all three scenarios we considered different values of n. Under each of the scenarios we generated $N=2000$ datasets, and for each dataset three different estimates were obtained, the MCL estimate, the JNF estimate, and the MDS estimate. In the JNF method we treated each stratum as an individual unit, and we estimated the parameters repeatedly by deleting one stratum at a time. For the purpose of comparisons, we presented the bias (with its empirical standard error), the estimated standard error (ESD), the 95 per cent coverage probability (CP) based on a Wald-type confidence interval, and the true (empirical) standard error TSD. In order to reduce the effect of some extreme observations we used the median absolute deviation (MAD): $\text{median}_{1\leqslant k\leqslant N}|\widehat{\beta}_k-\text{median}(\widehat{\boldsymbol{\beta}})|/0.6745$ for TSD, where $\widehat{\boldsymbol{\beta}}=(\widehat{\beta}_1,\widehat{\beta}_2,\ldots,\widehat{\beta}_N)^{\text{T}}$ and $\widehat{\beta}_k$ represents the estimate for the kth simulated dataset [12, p. 144]. In the simulation we deleted the datasets when the absolute value of the MCL or JNF estimates >10 (in such cases the corresponding standard errors were always >1000) and then calculated the summary quantities for all three approaches. We also presented results for the MDS method based on all 2000 replications in Tables I–III. Notice that the MDS estimates conditional on the datasets where both the MCL and JNF estimates were finite often show bias because a particular data configuration which may occur with non-zero probability was discarded. However, the performance of a method should be judged based on the collection of all datasets instead of its subset. The simulation results indicate that in all the scenarios considered the MDS method does not produce infinite values of the estimators. Therefore we should evaluate the MDS method based on all 2000 replications.

In the tables, the summary quantities for a method with an * refer to the approximated values obtained from the datasets where both MCL and JNF estimates are finite. In this context, the results in Tables I–III corresponding to scenarios 1–3, respectively, can be summarized as follows:

- The absolute bias of MCL estimator increases with β.
- For the simulation scenarios considered, the MDS estimator has less absolute bias than that of the JNF method for nonzero β and its estimates were always finite. However, the MCL and JNF estimates are infinite in many occasions, specially for $n=30$ or 50. For scenario 1 and when $n=30$ and $\beta=1.5$, about 50 per cent datasets yield infinite MCL estimate (results not shown here).
- The JNF method substantially reduces bias and the ESD of the estimate is always larger than its TSD.
- Overall, the variance of the MDS estimator is smaller than those of the other two estimators. We calculated ESD of the MDS estimator based on formula (6) which generally gives more accurate estimates of the TSDs (approximated

Table I. Results of the simulation study for one binary covariate and $M=1$.

	$n=30$				$n=50$				$n=100$			
Method	Bias (SE)	TSD†	ESD	CP	Bias (SE)	TSD†	ESD	CP	Bias (SE)	TSD†	ESD	CP
					$\beta=0.5$							
MCL*	−0.041 (0.016)	0.697	0.666	0.988	0.017 (0.011)	0.510	0.516	0.965	0.004 (0.008)	0.347	0.356	0.964
JNF*	−0.118 (0.014)	0.597	0.770	0.996	−0.030 (0.011)	0.475	0.561	0.986	−0.014 (0.008)	0.341	0.367	0.974
MDS*	−0.087 (0.015)	0.645	0.589	0.973	−0.017 (0.011)	0.479	0.478	0.957	−0.012 (0.008)	0.341	0.343	0.959
MDS	−0.001 (0.015)	0.670	0.599	0.945	−0.004 (0.011)	0.499	0.479	0.952	−0.012 (0.008)	0.341	0.343	0.959
					$\beta=1$							
MCL*	−0.110 (0.013)	0.529	0.658	0.975	0.018 (0.011)	0.505	0.524	0.977	0.024 (0.008)	0.349	0.364	0.955
JNF*	−0.248 (0.010)	0.418	0.772	0.985	−0.072 (0.010)	0.426	0.581	0.979	−0.012 (0.007)	0.330	0.380	0.975
MDS*	−0.194 (0.012)	0.499	0.578	0.958	−0.046 (0.010)	0.452	0.479	0.963	−0.007 (0.008)	0.336	0.349	0.951
MDS	−0.012 (0.013)	0.574	0.584	0.930	−0.014 (0.010)	0.443	0.482	0.951	−0.006 (0.008)	0.337	0.349	0.951

MCL, JNF, and MDS stand for the maximum conditional likelihood, jackknife, and modified score estimators, respectively.
TSD, ESD, and CP represent the empirical standard error, estimated standard error, and nominal 95 per cent coverage probability based on a Wald-type confidence interval of the point estimators.
SE represents the empirical standard error of bias.
*Estimate was calculated based on the datasets where both MCL and JNF are finite out of 2000 replications. When $\beta=0.5$ the number of deleted datasets out of 2000 replications is 165, 15, for $n=30$ and 50, respectively, and when $\beta=1$ the number of deleted datasets is 272, 40, 1 for $n=30$, 50, and 100, respectively.
†Approximation by the MAD method.

Table II. Results of the simulation study for one continuous covariate and $M=2$.

	$n=20$				$n=30$				$n=50$			
Method	Bias (SE)	TSD†	ESD	CP	Bias (SE)	TSD†	ESD	CP	Bias (SE)	TSD†	ESD	CP
					$\beta=0.5$							
MCL*	0.026 (0.017)	0.772	0.811	0.972	0.019 (0.014)	0.608	0.629	0.958	0.020 (0.010)	0.456	0.469	0.960
JNF*	0.008 (0.015)	0.691	0.949	0.981	0.008 (0.012)	0.559	0.701	0.972	0.012 (0.010)	0.425	0.499	0.967
MDS*	0.008 (0.015)	0.664	0.686	0.944	0.006 (0.012)	0.548	0.560	0.942	0.011 (0.010)	0.426	0.435	0.949
MDS	0.008 (0.015)	0.664	0.686	0.944	0.006 (0.012)	0.548	0.560	0.942	0.011 (0.010)	0.426	0.435	0.949
					$\beta=1$							
MCL*	0.078 (0.017)	0.759	0.803	0.967	0.056 (0.014)	0.622	0.629	0.961	0.038 (0.010)	0.460	0.471	0.956
JNF*	−0.038 (0.015)	0.654	0.935	0.977	−0.018 (0.012)	0.541	0.692	0.972	−0.005 (0.010)	0.435	0.496	0.971
MDS*	−0.014 (0.015)	0.665	0.675	0.932	−0.009 (0.013)	0.560	0.559	0.942	−0.002 (0.010)	0.436	0.438	0.940
MDS	−0.014 (0.015)	0.665	0.675	0.932	−0.009 (0.013)	0.560	0.559	0.942	−0.002 (0.010)	0.436	0.438	0.940
					$\beta=1.5$							
MCL*	0.191 (0.018)	0.790	0.851	0.973	0.126 (0.014)	0.638	0.666	0.966	0.081 (0.013)	0.492	0.499	0.958
JNF*	−0.028 (0.015)	0.686	0.983	0.982	−0.007 (0.013)	0.576	0.731	0.982	0.010 (0.011)	0.471	0.522	0.963
MDS*	0.019 (0.016)	0.696	0.699	0.927	0.013 (0.013)	0.591	0.586	0.929	0.015 (0.010)	0.465	0.461	0.936
MDS	0.019 (0.016)	0.696	0.699	0.927	0.013 (0.013)	0.591	0.586	0.929	0.015 (0.010)	0.465	0.461	0.936
					$\beta=2.0$							
MCL*	0.214 (0.019)	0.834	0.939	0.958	0.129 (0.015)	0.690	0.732	0.952	0.067 (0.012)	0.520	0.548	0.952
JNF*	−0.138 (0.016)	0.730	1.145	0.972	−0.059 (0.014)	0.634	0.816	0.956	−0.033 (0.011)	0.492	0.582	0.949
MDS*	−0.031 (0.016)	0.733	0.749	0.902	−0.026 (0.014)	0.631	0.635	0.914	−0.023 (0.011)	0.493	0.506	0.924
MDS	−0.028 (0.016)	0.734	0.751	0.902	−0.026 (0.014)	0.631	0.635	0.914	−0.023 (0.011)	0.493	0.506	0.924

MCL, JNF, and MDS stand for the maximum conditional likelihood, jackknife, and modified score estimators, respectively.
TSD, ESD, and CP represent the empirical standard error, estimated standard error, and nominal 95 per cent confidence interval coverage probability of the point estimators.
SE represents the empirical standard error of bias.
*Estimate was calculated based on the datasets where both MCL and JNF are finite out of 2000 replications. When $\beta=2$ the number of deleted dataset out of 2000 replication is 1, for $n=30$.
†Approximation by the MAD method.

Table III. Results of the simulation study for two binary covariates and $M=1$.

	$n=30$				$n=50$				$n=100$			
Method	Bias (SE)	TSD†	ESD	CP	Bias (SE)	TSD†	ESD	CP	Bias (SE)	TSD†	ESD	CP
					$\beta_1=0.5, \beta_2=0.5$							
MCL*	0.014 (0.015)	0.621	0.642	0.980	0.016 (0.011)	0.484	0.486	0.959	0.005 (0.008)	0.338	0.331	0.950
	−0.030 (0.015)	0.625	0.684	0.985	0.043 (0.012)	0.522	0.524	0.963	0.017 (0.008)	0.348	0.355	0.963
JNF*	−0.082 (0.012)	0.510	0.759	0.997	−0.032 (0.010)	0.443	0.531	0.985	−0.014 (0.007)	0.324	0.344	0.966
	−0.131 (0.012)	0.495	0.819	0.998	−0.018 (0.011)	0.473	0.582	0.989	−0.006 (0.007)	0.334	0.371	0.980
MDS*	−0.045 (0.013)	0.551	0.556	0.965	−0.021 (0.010)	0.450	0.446	0.950	−0.012 (0.007)	0.325	0.318	0.946
	−0.088 (0.013)	0.549	0.586	0.970	−0.001 (0.011)	0.482	0.476	0.952	−0.003 (0.008)	0.336	0.340	0.962
MDS	−0.001 (0.013)	0.582	0.565	0.949	−0.019 (0.010)	0.451	0.447	0.950	−0.012 (0.007)	0.325	0.318	0.946
	0.007 (0.014)	0.605	0.596	0.940	0.006 (0.011)	0.482	0.477	0.949	−0.003 (0.008)	0.336	0.340	0.962
					$\beta_1=1, \beta_2=1$							
MCL*	−0.061 (0.016)	0.608	0.670	0.972	0.062 (0.012)	0.532	0.532	0.967	0.037 (0.008)	0.354	0.362	0.960
	−0.081 (0.016)	0.605	0.699	0.977	0.073 (0.012)	0.523	0.564	0.976	0.065 (0.009)	0.396	0.387	0.960
JNF*	−0.239 (0.012)	0.468	0.817	0.988	−0.051 (0.011)	0.468	0.605	0.975	−0.008 (0.007)	0.332	0.382	0.972
	−0.269 (0.012)	0.464	0.858	0.989	−0.055 (0.010)	0.450	0.649	0.985	0.014 (0.008)	0.373	0.411	0.980
MDS*	−0.169 (0.014)	0.527	0.566	0.945	−0.019 (0.011)	0.485	0.475	0.952	−0.002 (0.008)	0.336	0.343	0.951
	−0.192 (0.014)	0.515	0.586	0.96	−0.015 (0.011)	0.473	0.499	0.961	0.022 (0.008)	0.377	0.365	0.952
MDS	0.002 (0.014)	0.636	0.585	0.916	0.001 (0.011)	0.491	0.478	0.943	−0.002 (0.008)	0.336	0.343	0.951
	0.057 (0.014)	0.647	0.606	0.907	0.028 (0.011)	0.492	0.505	0.948	0.022 (0.008)	0.377	0.365	0.952

MCL, JNF, and MDS stand for the maximum conditional likelihood, jackknife, and modified score estimators, respectively.
TSD, ESD, and CP represent the empirical standard error, estimated standard error, and nominal 95 per cent confidence interval coverage probability of the point estimators.
SE represents the empirical standard error of bias.
*Estimate was calculated based on the datasets where both MCL and JNF are finite out of 2000 replications. When $\beta_1=\beta_2=0.5$ the number of deleted datasets out of 2000 replications is 208 and 10, for $n=30$ and 50, respectively, and when $\beta_1=\beta_2=1$ the number of deleted datasets out of 2000 replications is 572 and 81, for $n=30$, 50, respectively.
†Approximation by the MAD method.

by TSD) than that based on $I(\widehat{\beta}_R)$. For example, in scenario 1 and when $\beta=1$, the ESD based on formula (6) and $I(\widehat{\beta}_R)$ are 0.584 and 0.699 for $n=30$, 0.482, and 0.582 for $n=50$, and 0.349 and 0.369 for $n=100$. The corresponding TSDs are given in Table I. Furthermore, the standard errors obtained from $I(\widehat{\beta}_R)$ are almost identical to the results obtained from $I(\widehat{\beta}_{MCL})$ which are presented in the tables as the ESD of the MCL estimator.

- The nominal 95 per cent CPs for MCL estimator are either close to 0.95 or higher although the MCL estimator has high bias. Further numerical investigation revealed that the MCL and its standard error are highly correlated. Therefore, for a large estimate of β, the standard error is also large, and thereby the confidence interval likely includes the true parameter.
- In some cases the MDS estimator has slightly lower CPs in its Wald-type confidence intervals. In unconditional logistic regression models, profile likelihood intervals have been shown to have better coverage properties than Wald-type intervals [8, 9, 13]. Here we computed penalized conditional-likelihood (PCL) based confidence intervals for the MDS estimator, and found that the corresponding CP is close to 0.95. For example, in scenario 2, for $\beta=2$ and $n=20$ and 30 the PCL-based CPs are 0.948 and 0.945, respectively. For small sample sizes, such as $n=30$ and 50, the CPs based on the PCL confidence interval appear to be better than those based on a Wald-type confidence interval. Overall, the simulation results indicate that Wald-type intervals for MDS (no matter what variance estimate) do not cover well when the bias preventive estimation is most desired.

Additional simulation study (not shown here) indicates that all three methods yield almost unbiased estimates when the true value of β is zero. For all three estimators in scenario 1 we also estimated the parameters by the method prescribed in Greenland [3]. For large values of β and for small sample size, Greenland's method reduces bias. For example, after deleting the datasets that MCL or JNF estimates are infinite, we found that when $\beta=2$ (not shown) and sample size $n=30$, the empirical bias due to Greenland and MCL estimators are 0.029 (0.731) and −0.725 (0.427) based on 1166 remaining datasets and for $n=50$ they are 0.131 (0.668) and −0.094 (0.487), based on 1673 remaining datasets, respectively. The quantity in the parentheses represents the empirical standard error of the estimate. It appears that the variance of the Greenland method is slightly larger than that of the MCL estimator. However, we did not find any appreciable differences between the Greenland's estimator and the MCL estimator for the sample sizes considered in the simulation when β is 0.5, 1, and 1.5 (not shown).

In summary, the limited simulation study along with the data example below illustrates the potential usefulness of the MDS estimator compared to the existing bias correction approaches in matched case–control studies.

Table IV. Results of the analysis of the $1:M$ matched case–control data on low birthweight study with two covariates, SMOKE and PTD. The JNF estimates are infinite when $M=1$.

		SMOKE			PTD		
M	Method	Estimate	SE	p-value	Estimate	SE	p-value
	MCL	0.598	0.476	0.209	1.733	0.611	0.005
3	JNF	0.587	0.520	0.259	1.597	0.705	0.023
	MDS	0.583	0.461	0.206	1.646	0.565	0.004
	MCL	0.699	0.615	0.256	1.896	1.083	0.080
1	MDS	0.636	0.547	0.245	1.542	0.740	0.037

'Estimate' and 'SE' denote the estimate and its standard error for the parameters of interest. Here 'SE' for MDS is based on the sandwich estimator defined in equation (4).

5. An analysis of low birthweight data

In order to illustrate the MDS method for $M_i=1$ and $M_i>1$ for all i we considered the 1:3 matched low birthweight dataset from Hosmer and Lemeshow [14] which were collected at the Baystate Medical Center, Springfield, MA, in 1986. Since here all M_i are the same, without any ambiguity we will use M for M_i. Low birthweight is one of the main concerns to physicians, and possibly one of the important visible causes of infant deaths. Scientists believe that the mother's behavior during pregnancy (prenatal care, smoking, diet, etc.) can play a major role for her baby's birthweight. For this dataset an infant was defined as a case if its birthweight was below 2500 g, otherwise the infant was a control. Note that for the dataset mother's age was between 16 and 32 years. Among several covariates, we focused on two covariates, the mother's smoking status (SMOKE) during the pregnancy and presence of previous preterm delivery (PTD).

First we considered 1:3 matching, and analyzed the data by the MCL method, the JNF method, and the MDS approach. The results are presented in the upper half of Table IV. Note that the p-values are based on the asymptotic distribution of the statistics. It is seen that at level 5 per cent all three estimators MCL, JNF, and MDS are statistically significant for PTD, while there is no significant association between SMOKE and the risk of having a low birthweight child.

For the example of using $M=1$, we randomly picked one control out of 3 from each stratum of the dataset and formed a 1:1 matched case–control data. The results of the analysis of the 1:1 matched data are presented at the bottom half of Table IV. Here, we also considered SMOKE and PTD as the two covariates. Note that in this scenario the JNF estimate is infinite. The MDS estimate of the log-odds ratio for PTD is statistically significant at level 5 per cent while the corresponding MCL estimate is not. Both the MDS and MCL estimates of the log-odds ratio parameter for SMOKE turn out to be statistically insignificant at level 5 per cent.

The table also provides the p-values which are calculated based on the asymptotic Z-statistics.

6. Discussion

In this paper we apply Firth's general approach to reduce the bias in the MCL estimator for matched case–control studies. The MDS estimator is obtained as the solution of a modified conditional score equation. Numerical studies show that the MDS approach not only reduces bias but also generally has less variance than the MCL estimators. The MDS technique seems to yield finite parameter estimates, and can be applied when MCL and JNF estimates are infinite. Another advantage of the MDS approach over the JNF method is that the computation is easy and much less time consuming. Furthermore, like other methods that improve the bias, the MDS method can also handle multiple covariates simultaneously, and the covariates could be categorical, continuous, or a mixture of both types.

It appears possible to apply the MDS method to obtain MDS estimates of the log-odds ratio parameters when a covariate is partially missing in the dataset. In this situation the likelihood will be more complex. A main issue in that context is how to appropriately calculate the information matrix. This problem deserves further investigation and is a topic for future research.

For the conditional logistic regression analysis we used `clogit` function of the statistical software `R`, and for the MDS estimator we used the Newton–Raphson method. The computer code is available at `http://www.stat.tamu.edu/~sinha/research.html`. Since Heinze and Schemper [9] adopted Firth's approach to handle the issue of monotonicity in Cox's partial likelihood and conditional logistic regression is a special case of Cox's partial likelihood for a stratified survival design, it is conceivable that one could adopt the software due to Heinze and Schemper [9] to obtain the MDS estimates.

Appendix A: Expression for the terms used in equation (4)

Note that

$$\frac{\partial U_{(r)}^{\text{mod}}(\boldsymbol{\beta})}{\partial \beta_k} = \frac{\partial U_{(r)}(\boldsymbol{\beta})}{\partial \beta_k} + \frac{1}{2} tr \left\{ -I^{-1}(\boldsymbol{\beta}) \frac{\partial I(\boldsymbol{\beta})}{\partial \beta_k} I^{-1}(\boldsymbol{\beta}) \frac{\partial I(\boldsymbol{\beta})}{\partial \beta_r} + I^{-1}(\boldsymbol{\beta}) \frac{\partial^2 I(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_r} \right\},$$

where $\partial I(\boldsymbol{\beta})/\partial \beta_r = \sum_{i=1}^{n} \sum_{j=1}^{M_i+1} (\boldsymbol{X}_{ij} - 2\bar{\boldsymbol{X}}_{i\cdot}) \boldsymbol{X}_{ij}^{\top} (X_{ijr} - \bar{X}_{i\cdot r}) p_{ij}$ and

$$\frac{\partial^2 I(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_r} = \sum_{i=1}^{n} \sum_{j=1}^{M_i+1} \left[\left(-2 \frac{\partial \bar{\boldsymbol{X}}_{i\cdot}}{\partial \beta_k} \boldsymbol{X}_{ij}^{\top} \right) (X_{ijr} - \bar{X}_{i\cdot r}) p_{ij} + (\boldsymbol{X}_{ij} - 2\bar{\boldsymbol{X}}_{i\cdot}) \boldsymbol{X}_{ij}^{\top} \left\{ (X_{ijr} - \bar{X}_{i\cdot r}) \frac{\partial p_{ij}}{\beta_k} - \frac{\partial \bar{X}_{i\cdot r}}{\beta_k} p_{ij} \right\} \right],$$

with $\partial \bar{\boldsymbol{X}}_{i\cdot}/\partial \beta_k = \sum_{j=1}^{M_i+1} \boldsymbol{X}_{ij} (\partial p_{ij}/\partial \beta_k)$, $\partial \bar{X}_{i\cdot r}/\partial \beta_k = \sum_{j=1}^{M_i+1} X_{ijr} (\partial p_{ij}/\partial \beta_k)$, and $\partial p_{ij}/\partial \beta_k = p_{ij}(X_{ijk} - \bar{X}_{i\cdot k})$.

References

1. Jewell NP. Small-sample bias of point estimators of the odds ratio from matched sets. *Biometrics* 1984; **40**:421–435.
2. Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analyses*: *Theory and Practice*. MIT Press: Cambridge, MA, 1975.
3. Greenland S. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. *Biostatistics* 2000; **1**:113–122.
4. Cordeiro GM, McCullagh P. Bias correction in generalized linear models. *Journal of the Royal Statistical Society*, *Series B* 1991; **53**:629–643.
5. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993; **80**:27–38.
6. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Statistics in Medicine* 2002; **21**:2409–2419.
7. Bull SB, Mak C, Greenwood CMT. A modified score function estimator for multinomial logistic regression in small samples. *Computational Statistics and Data Analysis* 2002; **39**:57–74.
8. Bull SB, Lewinger JP, Lee SSF. Confidence intervals for multinomial logistic regression in sparse data. *Statistics in Medicine* 2007; **26**:903–918.
9. Heinze G, Schemper M. A solution to the problem of monotone likelihood in Cox regression. *Biometrics* 2001; **57**:114–119.
10. Breslow NE, Day NE. *Statistical Methods in Cancer Research*, *Volume 1—The Analysis of Case–Control Studies*. IARC: Lyon, 1980.
11. Cox DR, Snell EJ. *Analysis of Binary Data* (2nd edn). Chapman & Hall: London, 1968.
12. Huber PJ. *Robust Statistics*. Wiley: New York, 1981.
13. Heinze G. A comparative investigation of methods for logistic regression with separated or nearly separated data. *Statistics in Medicine* 2006; **25**:4216–4226.
14. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. Wiley: New York, 1989.