STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:3240–3257
Published online 5 January 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2790

Analysis of matched case–control data with multiple ordered disease states: Possible choices and comparisons

Bhramar Mukherjee[1,*,†], Ivy Liu[2] and Samiran Sinha[3]

[1]*Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[2]*School of Mathematics, Statistics, and Computer Science, Victoria University of Wellington, Wellington, New Zealand*
[3]*Department of Statistics, Texas A and M University, College Station, Texas 77843, U.S.A.*

SUMMARY

In an individually matched case–control study, effects of potential risk factors are ascertained through conditional logistic regression (CLR). Extension of CLR to situations with multiple disease or reference categories has been made through polychotomous CLR and is shown to be more efficient than carrying out separate CLRs for each subgroup. In this paper, we consider matched case–control studies where there is one control group, but there are multiple disease states with a natural ordering among themselves. This scenario can be observed when the cases can be further classified in terms of the seriousness or progression of the disease, for example, according to different stages of cancer. We explore several popular models for ordered categorical data in this context. We first adopt a cumulative logit or equivalently, a proportional-odds model to account for the ordinal nature of the data. The important distinction of this model from a stratified dichotomous and polychotomous logistic regression model is that the stratum-specific nuisance parameters cannot be eliminated in this model *via* the conditional-likelihood approach. We discuss a Mantel–Haenszel approach for analysing such data. We point out possible difficulties with standard likelihood-based approaches with the cumulative logit model when applied to case–control data. We then consider an alternative conditional adjacent-category logit model. We illustrate the methods by analysing data from a matched case–control study on low birthweight in newborns where infants are classified according to *low* and *very low* birthweight and a child with normal birthweight serves as a control. A simulation study compares the different ordinal methods with methods ignoring sub-classification of the ordered disease states. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: ordinal response; proportional odds; matched case–control; adjacent category logit; conditional likelihood

[*] Correspondence to: Bhramar Mukherjee, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

[†] E-mail: bhramar@umich.edu, mukherjee@stat.ufl.edu

Contract/grant sponsor: NSA young investigator; contract/grant number: H98230-06-1-0033

Received 12 September 2005
Accepted 3 November 2006

1. INTRODUCTION

Case–control studies are important tools in investigating the aetiology of rare diseases like cancer. The fundamental goal is to explore association between the disease and its potential risk factors. In modern medicine, with precise characterization of disease states in histological and morphological terms, it is natural to note that the disease state might have more than one category, i.e. we may have subdivisions within the 'cases'. For example, patients diagnosed with cancer may have cancer of stage-I, stage-II or stage-III at the time of the diagnosis which is an example of ordinal disease categories. There are several popular models for analysing ordinal response [1] which may be readily adapted to analyse case–control data with ordered disease categories. However, one must be careful that the chosen prospective ordinal model leads to consistent estimation of the covariate-response association parameters of interest, as will be obtained from the appropriate retrospective likelihood generated by the case–control sampling design.

Matching is often implemented in a case–control design in order to avoid bias due to potential confounders. A proper statistical analysis should account for the matched design. Unmatched analysis of matched data, ignoring the effect of matching, generally yields conservative estimates of the relative risk [2]. Breslow *et al.* [3] proposed the well-known conditional logistic regression (CLR) method for analysing matched case–control data. Usual unconditional maximum-likelihood (ML) inference for such finely stratified data with large number of nuisance parameters are potentially subject to the Neyman–Scott phenomenon. Conditioning on the complete sufficient statistic for the matched set specific parameters in each matched set eliminates these nuisance parameters and the score equation based on the conditional likelihood provides the optimum estimating function [4].

Liang and Stewart [5] extended the usual CLR methodology to polychotomous disease or reference categories. They established that separating the disease or reference states into subgroups and conducting CLR pairwise is less efficient than the polychotomous conditional logistic regression (PCLR) approach. Liang and Stewart [5], Becher and Jöckel [6], Becher [7], apply their methods to matched case–control studies with two control groups, typically the hospital controls and the population controls. Thomas *et al.* [8] and Durbin and Pasternack [9] apply polychotomus logistic regression model to analyse data with multiple disease groups and one set of controls. Recently, Sinha *et al.* [10] considered a Bayesian semiparametric model for analysing matched case–control data with multiple disease states and missingness in exposure values.

None of the above papers consider the situation when the underlying probability model for the disease incidence acknowledges a possible natural ordering of the disease states when it is present. They all incorporate the multinomial logistic regression model as the underlying disease-risk model. The most commonly used model for ordinal logistic regression is the cumulative logit or the proportional odds model. One major difference between the multinomial logit model and the cumulative logit model when applied to matched samples is that in the latter, there are no sufficient statistics for the stratum specific nuisance parameters, so the usual conditioning technique to eliminate the nuisance parameters does not apply. In absence of any reduction due to sufficiency, for matched pair data with ordinal response, McCullagh [11] suggested an ingenious approach for obtaining consistent estimates. Agresti and Lang [12] propose fitting simultaneous conditional ML estimates with all possible binary collapsing of the ordinal response to eliminate the nuisance parameters in order to achieve consistent estimation.

In general, there could be several approaches to analyse highly stratified ordinal data using the proportional odds model. Liu and Agresti [13] propose a Mantel–Haenszel (MH) type estimate for

DOI: 10.1002/sim

the odds ratio (OR) in a cumulative logit model which can be adapted to 1:M matched case–control study with ordinal disease states and a single categorical exposure. Another possible alternative which can handle both categorical and continuous exposure is to posit the problem in a Bayesian paradigm, and assume suitable priors on all the parameters and conduct posterior inference *via* implementing a Markov chain Monte Carlo numerical integration technique. Alternatively, one can propose a random effects model (REM) for stratification parameters akin to Hedeker and Gibbons [14]. There are choices beyond the proportional odds model: the adjacent category logit model for ordered data [1], the PCLR analysis ignoring the natural ordering of disease states, the CLR analysis by collapsing the disease states into a single category. The purpose of this article is to explore, illustrate, compare and contrast the usage and performance of these possible choices for analysing matched case–control data.

The rest of the paper is organized as follows. Section 2 considers possible choices with the stratified proportional odds model. Section 2.1 considers MH estimation in stratified contingency tables with ordinal response. Section 2.2 considers a natural choice for handling stratum specific nuisance parameters in a full Bayesian framework (Section 2.2.1). In Section 2.2.2, we discuss the problems with the use of Bayesian method using the prospective cumulative logit link for individually matched case–control data, specifically: (i) the failure to acknowledge the retrospective sampling design *via* the prospective likelihood and (ii) selection of priors on the nuisance parameters in finely stratified sparse data situations. In Section 3, we consider the conditional adjacent category logit model. In Section 4, we describe the CLR and polychotomous CLR models which do not account for the ordinal nature of the response. The analysis of a real data set by various methods is considered in Section 5. Section 6 presents the results of a simulation study. Section 7 contains concluding remarks and final recommendations.

2. THE STRATIFIED PROPORTIONAL ODDS MODEL

Let us first describe the general data structure used throughout the paper. We consider a 1:M matched case–control study with n matched sets. Let Y_{ij}, $i=1,\ldots,n$, $j=1,\ldots,M+1$, denote the disease status (for our example, the birthweight category) of subject j in matched set or stratum i. Also let Y_{ij} be a c-category ordinal variable with categories scaled from $1,\ldots,c$, with reference category c denoting the control group and the subdivisions within the diseased group denoted by category $1,\ldots,c-1$. Let X_{ij} denote the observed categorical exposure observed for subject j in matched set i. For our example on low-birthweight data in newborns, $Y_{ij}=3$ denotes the control group (normal birth weight >2500 g), $Y_{ij}=2$ denotes disease group 2 (low birthweight between 2000 and 2500 g), and $Y_{ij}=1$ denotes disease group 1 (very low birthweight <2000 g) whereas X_{ij} is a binary exposure, denoting smoking status of the mother.

The most popular model for ordinal data was inspired by McCullagh [15] by modelling the log odds corresponding to the cumulative probabilities. The *cumulative logit* model for disease incidence is given by

$$\text{logit}[P(Y_{ij}\leqslant k|X_{ij})]=\alpha_k+\gamma_i+\beta X_{ij}, \quad i=1,\ldots,n,\ j=1,\ldots,M+1, k=1,\ldots,c-1 \quad (1)$$

where $\alpha_1<\alpha_2<\cdots<\alpha_{c-1}$ and $P(Y_{ij}\leqslant c)=1$. The parameters $\{\alpha_k\}$, called *cut points*, are usually nuisance parameters of little interest. The parameters $\{\gamma_i\}$ are stratum-specific nuisance parameters; the total number of these nuisance parameters, γ_i, increases with sample size. This model applies simultaneously to all $c-1$ cumulative probabilities, and it assumes an identical effect of the

DOI: 10.1002/sim

predictors for each cumulative probability. This particular type of cumulative logit model, with exposure effect β the same for all k, is often referred to as a *proportional odds model* [15]. For example, let $X_{ij}=1$ for a smoker mother and $X_{ij}=0$ for a non-smoker mother. The model assumes that the odds that Y_{ij} falls below level k for a smoker mother are $\exp(\beta)$ times the odds for a non-smoker mother, for $k=1,2$. We refer to $\theta=\exp(\beta)$ as the *cumulative OR* for the conditional association of weight of newborn and smoking status of mother.

For a case–control study with binary response ($c=2$), the model in (1) is simply a logistic regression model. When the model holds, but n is large and the data are sparse, the ML estimator of β tends to overestimate the true log OR and leads to biased and inconsistent estimation. This is a phenomenon observed in models where the number of parameters (such as $\{\gamma_i\}$ in (1)) grows at the same rate as the sample size [16]. For instance, for a 1:1 pair matched case–control study, the ML estimator of β converges to double the true value [17, p. 244]. Similar phenomenon occurs for the proportional odds model (1) when $c>2$ [13]. As mentioned in the Introduction, for $c>2$, conditional ML method cannot be applied to the proportional odds model (1). In the following two subsections we discuss possible estimation strategies with underlying probability model (1).

2.1. Mantel–Haenszel estimation

Classical methods for making inference on the OR for pair-matched data with a dichotomous exposure are based on conditioning on the marginal totals of the resultant 2×2 tables and uses only the *discordant* pairs. Tests for hypothesis of no association between the disease and exposure in matched samples were proposed by McNemar [18]. The seminal paper by Mantel and Haenszel [19] furnished the MH estimate of the summary relative risk for stratified data and showed a chi-squared test of conditional independence. Mantel [20] generalized the chi-squared test in a series of $r\times c$ tables for ordinal variables. There have been several extensions of MH methods to situations when one has 1:M matching or variable matching ratios, and for categorical exposure [21]. Based on the proportional odds model (1), Liu and Agresti [13] proposed a MH-type estimator of an assumed common cumulative OR ($\exp(\beta)$), which is an extension of the ordinary MH estimator. Their estimator has behaviour similar to the MH estimator of a common OR for several 2×2 tables. It is consistent even when the data are sparse, that is, even when the number of matched sets (strata) increases proportional to the sample size, which is the case in an individually matched case–control study. The method is described below.

For a matched case–control study with c ordered disease categories and a binary exposure, we can cross-classify each matched stratum into a $2\times c$ table, where the row variable is the exposure and the column variable is the disease. In total, there are n matched strata. Therefore, there are n such $2\times c$ tables. We use notations Z_{rki} to denote the cell counts for the row r, column k, and stratum i, where $r=1,2$, $k=1,\ldots,c$, and $i=1,\ldots,n$. Let n_{ri} be the total number of subjects in row r and stratum i. Let us also denote the cumulative counts in row r and stratum i by $Z^*_{rki}=Z_{r1i}+\cdots+Z_{rki}$. Also, let the total number of subjects in the ith stratum be denoted by $N_i=\sum_r\sum_k Z_{rki}$. Then Liu–Agresti's MH-type estimator equals

$$\hat{\theta}=\frac{\sum_{i=1}^{n}\sum_{k=1}^{c-1} Z^*_{1ki}(n_{2i}-Z^*_{2ki})/N_i}{\sum_{i=1}^{n}\sum_{k=1}^{c-1}(n_{1i}-Z^*_{1ki})Z^*_{2ki}/N_i} \tag{2}$$

For model (1), the same cumulative OR occurs for all collapsings of the ordinal disease categories into the binary category ($\leqslant k$, $>k$), $k=1,\ldots,c-1$. Suppose we naively treat $c-1$ different 2×2 collapsed tables of each stratum as independent. Estimator (2) is simply the ordinary MH estimator

DOI: 10.1002/sim

of the common OR for $n(c-1)$ separate 2×2 tables. However, the variance estimator of $\hat{\theta}$ needs to take the dependency of the collapsing tables for each stratum into account. Liu and Agresti [13] proposed the variance estimator as

$$\widehat{\mathrm{Var}}[\log(\hat{\theta})] = \frac{\sum_{i=1}^{n} \hat{\xi}_i(\hat{\theta})}{\hat{\theta}^2 (\sum_{i=1}^{n} \sum_{k=1}^{c-1} (n_{1i} - Z^*_{1ki}) Z^*_{2ki}/N_i)^2}$$

where $\hat{\xi}_i(\theta) = \sum_{k=1}^{c-1} \hat{\phi}_{kki}(\theta) + 2\sum_{k<k'}^{c-1} \hat{\phi}_{kk'i}(\theta)$ with

$$\hat{\phi}_{kk'i}(\theta) = \frac{n_{1i}n_{2i}}{N_i^2}\left\{\frac{\theta(n_{1i}-Z^*_{1k'i})Z^*_{2ki}}{n_{1i}}\left[1+(\theta-1)\frac{Z^*_{2k'i}}{n_{2i}}\right]\right.$$
$$\left. + \frac{Z^*_{1ki}(n_{2i}-Z^*_{2k'i})}{n_{2i}}\left[\theta-(\theta-1)\frac{Z^*_{1k'i}}{n_{1i}}\right]\right\}, \quad k\leqslant k' = 1,\ldots,c-1$$

The proof of consistency in a large strata, sparse data set-up was given by Liu and Agresti [13]. This MH-type estimator also applies to a proportional odds model for a more general setting with $\alpha_k + \gamma_i$ in (1) replaced by matched set specific cut-off points α_{ki}.

For a 1:1 matched case–control study, the MH-type estimator (2) simplifies dramatically. Consider only the matched pairs with 'discordant' exposures, that is, where the case and control subject fall into two different exposure categories. We can express the Liu–Agresti MH estimator in terms of counts in a $c\times c$ table that represents the joint disease classification for such pairs with discordant exposure information. The data sets for the 1:1 and 1:3 matched studies are available at http://www.sph.umich.edu/bhramar/public_html/research. For instance, in the 1:1 matched low-birthweight data, matched pair number 11 has different exposure categories for case and control subjects and the joint disease categories are 1 (very low birthweight for the smoker mother), and 3 (normal birthweight for the non-smoker mother). Therefore, we count this as an observation in row 1 and column 3 in the 3×3 table ($c=3$ for this example). In total, there are 7 such pairs. So the ultimate cell count in cell(1, 3) of the 3×3 table is 7. There are only 30 matched pairs out of the 56 matched pairs which have discordant exposure information (different smoking habits). The summary 3×3 table with row variable as disease status of a child with a smoker mother, column variable as disease status of a child with a non-smoker mother, and the cell counts as how many matched pairs fall in each joint disease classification cell is displayed below. Note that no counts are available for many of the cells because of the intrinsic structure of the matched case–control study design.

		Birthweight category of a child with non-smoker mother		
		1	2	3
Birthweight	1	—	—	7
category of a child	2	—	—	15
with smoker mother	3	3	5	—

In general, let $m_{kk'}$ be the sample count for row k and column k' in the $c \times c$ table. The Liu–Agresti MH estimate of θ can be expressed as

$$\hat{\theta} = \frac{\sum_{k<k'} (k'-k) m_{kk'}}{\sum_{k>k'} (k-k') m_{kk'}} \tag{3}$$

Agresti and Lang [12] also proposed this estimator for a proportional odds model with matched pairs. They derive this estimator by simultaneously fitting CLR estimates to all possible binary collapsing of the response categories and also establish its consistency properties.

When the subjects in a matched stratum all have non-smoker mother or all have smoker mother, that is, there is a zero count either for n_{1k} or n_{2k}, the information from that stratum does not contribute to the MH-type estimator (2). This is not surprising, as for the usual case–control study with binary response, when one uses the ordinary MH estimator to describe the conditional association between disease and exposure, only the discordant pairs are used. Therefore, there is no significant loss of information regarding β, by ignoring those strata. For pair-matched case–control data one can simply use the simplified version of the MH estimator (3). For the 1:1 matched data set with 56 strata, as represented in the table above, we have $m_{13}=7$, $m_{23}=15$, $m_{31}=3$, $m_{32}=5$, and zeros otherwise. The estimated OR as computed by the Agresti–Lang recipe (3) is $\hat{\theta}=(15\cdot 1+7\cdot 2)/(5\cdot 1+3\cdot 2)=2.64$ which is identical with the Liu–Agresti estimate (2).

2.2. Likelihood-based approach

To obtain the likelihood function we construct the usual product of multinomials. Without loss of generality, we assume that the first subject in each stratum is a case and the rest are controls (i.e. $Y_{ij}=c$, for $j=2,\ldots,M+1$). The likelihood for the proportional odds model (1) has the form

$$\begin{aligned} L(\beta,\alpha_k,\gamma_i) &= \prod_{i=1}^{n} \prod_{j=1}^{M+1} \left[\prod_{k=1}^{c} (P(Y_{ij}\leqslant k|X_{ij}) - P(Y_{ij}\leqslant k-1|X_{ij}))^{I[Y_{ij}=k]} \right] \\ &= \prod_{i=1}^{n} \left[\prod_{k=1}^{c-1} \left(\frac{\exp(\alpha_k+\gamma_i+\beta X_{i1})}{1+\exp(\alpha_k+\gamma_i+\beta X_{i1})} - \frac{\exp(\alpha_{k-1}+\gamma_i+\beta X_{i1})}{1+\exp(\alpha_{k-1}+\gamma_i+\beta X_{i1})} \right)^{I[Y_{i1}=k]} \right. \\ &\quad \left. \times \prod_{j=2}^{M+1} \left(1 - \frac{\exp(\alpha_{c-1}+\gamma_i+\beta X_{ij})}{1+\exp(\alpha_{c-1}+\gamma_i+\beta X_{ij})} \right) \right] \end{aligned} \tag{4}$$

To handle the large number of nuisance parameters in (4), the two popular approaches are (i) use a Bayesian approach with a prior structure on γ_i and (ii) to use random effects terms to describe the strata effects. As we discuss in Section 2.2.2, these likelihood-based methods could potentially lead to biased estimation of the association parameter, due to lack of conformation with the retrospective nature of the case–control sampling design and also due to the sparsity of the data. We also noted in our research that the issues with the REM are similar to the full Bayesian method with a flat prior on β, thus we present results with only the Bayesian analysis. One can find details of the REM method and codes at http://www.sph.umich.edu/bhramar/public_html/research.

DOI: 10.1002/sim

2.2.1. A full Bayesian analysis. In a Bayesian paradigm, starting with the likelihood (4) of the data, one could assume suitable proper priors on β and the parameters γ_i and α_k, and proceed to estimate β by the mean of the posterior distribution of β given the data and all other parameters.

The model requires $\{\alpha_k\}$ to be increasing in k, as $P(Y_{ij}\leqslant k)$ is an increasing function of k. To ensure this we reparameterize the model in the following way: we set $\lambda_k = \log(\alpha_k - \alpha_{k-1})$, $k=2,\ldots,c-1$ with $\lambda_1=\alpha_1=0$. Thus, $\alpha_k=\sum_{i=1}^{k}\exp(\lambda_i)$, and λ_k are allowed to vary freely. One may choose suitable priors on λ_k. Alternatively, in a Bayesian paradigm one could put a positive prior on the successive difference of the cut-off values $\alpha_k - \alpha_{k-1}$. Specifically, we assumed the following parametric prior structure:

$$\alpha_k - \alpha_{k-1} \sim \text{Gamma}(a, b)$$

$$\gamma_i \overset{\text{iid}}{\sim} \text{N}(\mu_\gamma, \sigma_\gamma^2), \quad i=1,\ldots,n$$

$$\beta \sim \text{N}(\mu_\beta, \sigma_\beta^2)$$

We will call this model to be PBV standing for a fully parametric Bayes model with varying stratum effects.

Posterior inference is carried out by computing the posterior mean and highest posterior density credible interval. We use Gibbs sampler technique [22] to generate random numbers from the full conditional distribution of each parameter given all other parameters and data. We skip fairly standard algorithmic details for the sake of brevity.

Remark 1

We could also consider a simpler model with a much reduced number of parameters, simply by assuming a constant stratum effect model, $\gamma_i \equiv \gamma_0 \sim \text{N}(\mu_\gamma, \sigma_\gamma^2)$. We could then impose a normal prior on γ_0. But with a constant stratum effect model, which is essentially carrying out an unmatched analysis for matched data, it is easy to obtain the maximum likelihood estimate (MLE) by running any standard software. We will see in our data analysis that assuming a constant stratum effect model will lead to conservative estimate of the association parameter.

Remark 2

For the PBV model, one could question the i.i.d. parametric assumption on the prior distribution of γ_i. The parametric assumption could be relaxed by modelling the distribution of γ_i non-parametrically as done in Sinha *et al.* [10] which allows to data adaptively determine the degree of stratification using a Dirichlet Process prior on the distribution of γ_i. Since we want to focus on simple, ready to use methods in this paper we refrain from adopting computationally complex non-parametric models for the stratum effects.

Also, we would like to emphasize that one must be careful in choosing the prior distribution on the nuisance parameters in this sparse data scenario to avoid the risk of inconsistent estimation. For binary matched pair data with $c=2$, apparently innocuous choices for the Bayesian prior on the nuisance parameters can lead to extremely poor, exponentially inconsistent estimates of disease–exposure association [23]. Characterization results for a sensible class of priors have been provided by Rice [24] for $c=2$ with binary exposure. Our research indicates, for $c>2$, the Bayes estimates do reflect a certain degree of bias, and are sensitive to the choice of priors, and thus more careful thoughts may be necessary in characterizing the class of priors that produce consistent estimates of β.

2.2.2. Prospective proportional odds model and retrospective sampling scheme. The PBV method described above does not recognize the retrospective nature of the study design and use the fully prospective likelihood in (4) which will be obtained in a cohort study. The appropriate retrospective likelihood reflecting the sampling scheme for case–control studies with further disease sub-classification can be viewed as follows. We refer to the sampling scheme followed in the actual low-birthweight data set. At the first stage of sampling, the disease category (D) is recorded (case/control) and we then record the exposure values $X|D=d$. After recording this information, using some external source of information (in this case the actual birthweight) we ascertain the sub-group to which the diseased subjects (cases) belong, namely very low ($Y=1$) and low ($Y=2$). Let k_i denote the case status in matched set i, in our example it could be 1 or 2. Following the above sampling scheme, the appropriate likelihood thus would be

$$
\begin{aligned}
L_R &= \prod_{i=1}^{n} P(Y_{i1}|D_{i1}=1, X_{i1}) P(X_{i1}|D_{i1}=1) \times \prod_{i=1}^{n} \prod_{j=2}^{M+1} P(X_{ij}|D_{ij}=0) \\
&= \prod_{i=1}^{n} \frac{P(Y_{i1}, D_{i1}=1, X_{i1})}{P(D_{i1}=1)} \times \prod_{i=1}^{n} \prod_{j=2}^{M+1} P(X_{ij}|Y_{ij}=c) \\
&= \prod_{i=1}^{n} \frac{P(Y_{i1}=k_i, X_{i1})}{P(D_{i1}=1)} \times \prod_{i=1}^{n} \prod_{j=2}^{M+1} P(X_{ij}|Y_{ij}=c) \\
&= \prod_{i=1}^{n} P(X_{i1}|Y_{i1}=k_i) \frac{P(Y_{i1}=k_i)}{\sum_{r=1}^{c-1} P(Y_{i1}=r)} \times \prod_{i=1}^{n} \prod_{j=2}^{M+1} P(X_{ij}|Y_{ij}=c) \qquad (5)
\end{aligned}
$$

Prentice and Pyke [25] established that if the prospective model is logistic ($c=2$), then the MLE of the OR parameter β from the prospective likelihood is the same as the MLE obtained from the retrospective likelihood with correct asymptotic variance. For $c=2$, it has been noted by other authors [26] that ignoring this prospective–retrospective equivalence and using other links like the probit or log–log link produces biased estimate of the parameter of interest β. Unfortunately, the estimate of β obtained from the prospective likelihood induced by the cumulative logit link and the retrospective likelihood is not the same. Scott and Wild [27] propose a method to fit prospective models with any link function under outcome-dependent sampling schemes which requires knowledge of supplementary information on the population totals for cases and controls in each stratum which is not the situation we consider in the current paper.

The intuitive reason for this lack of agreement is fairly clear. For $c=2$, among all the generalized linear models for binary data, the logistic link function is the only one which absorbs the difference in sampling rates for each disease category in the intercept term as an offset and the covariate effects remain unaltered [28]. For multiple disease categories, the same phenomenon is replicated and the effect of the sampling rates for each disease category can be absorbed in the intercept parameter if and only if one uses a link with a multiplicative intercept structure, otherwise the sampling rates will affect estimation of the OR parameters as well. The degree of this design bias with the cumulative logit link will depend on the sampling ratio of different case subgroups to controls. We thank a referee for drawing our attention to this issue.

Remark 3
One may naturally think about direct retrospective modelling using (5), as partly done by Sinha *et al.* [10]. One can start with a model for the control distribution of the exposure $p(X|Y=c)$. As done by Sinha *et al.* [10], the retrospective likelihood L_R in (5) can be explicitly written in terms of the disease risk model as given in (1) and a model for control distribution of the exposure. As with any retrospective model, it may be quite hard to pose a 'robust' parameteric model for the multivariate exposure distribution and to adequately capture latent stratification effects. A prospective model is much easier to formulate and implement. However, under this sampling scheme, one must keep in mind that the likelihood-based methods using prospective cumulative logit model, may not be giving the true answer but only provide a precursor to the real association.

3. CONDITIONAL ADJACENT CATEGORY LOGIT MODEL

The adjacent category model [1] offers another possibility to model ordered response. The model is given by

$$\log \frac{P(Y_{ij}=k)}{P(Y_{ij}=k+1)} = \gamma_i + \alpha_k + \beta X_{ij}, \quad i=1,\ldots,n, \quad j=1,\ldots,M+1, \quad k=1,\ldots,c-1 \tag{6}$$

It is possible to make the adjacent category log odds parameter β to depend on k by introducing additional parameters, but we consider the constant effect model for simplicity. For case–control data, a varying β may be more meaningful as the log odds of control to a case category is often different than the log-odds of two successive-case subclasses. The model in (6) can be alternatively expressed in the multinomial logit structure as

$$\begin{aligned}\log \frac{P(Y_{ij}=k)}{P(Y_{ij}=c)} &= (c-k)\gamma_i + \sum_{r=1}^{c-k} \alpha_r + (c-k)\beta X_{ij} \\ &= \gamma^*_{ik} + \beta^*_k X_{ij}, \quad k=1,\ldots,c-1\end{aligned}$$

Where $\gamma^*_{ik} = (c-k)\gamma_i + \sum_{r=1}^{c-k} \alpha_r$ and $\beta^*_k = (c-k)\beta$. Let us, as before, assume without loss of generality that the first subject in each matched set is a case and the rest are controls. Let k_i again denote the disease status of the case subject in matched set i, $k_i \in (1,\ldots,c-1)$. To eliminate the nuisance parameters γ^*_{ik}, one could appeal to the conditional-likelihood principle and condition on the event $\sum_{j=1}^{M+1} Y_{ij} = k_i$ in matched set i. The conditional likelihood can be derived through the following steps.

Note that the above model states that the conditional probabilities of the disease variable given the exposure and the stratum are given by

$$P(Y_{ij}=k|X_{ij}) = \frac{\exp\{\gamma^*_{ik} + \beta^*_k X_{ij}\}}{1 + \sum_{r=1}^{c-1} \exp\{\gamma^*_{ir} + \beta^*_r X_{ij}\}} \quad \text{for } k=1,\ldots,c-1 \tag{7}$$

DOI: 10.1002/sim

and

$$P(Y_{ij}=c|X_{ij})=\frac{1}{1+\sum_{r=1}^{c-1}\exp\{\gamma_{ir}^{*}+\beta_{r}^{*}X_{ij}\}} \tag{8}$$

The conditional likelihood is simply

$$\begin{aligned}L_{\text{CADJ}}&=\prod_{i=1}^{n}P\left(Y_{i1}=k_i, Y_{ij}=c, j=2,\ldots,M+1|X_{ij}, \sum_{j=1}^{M+1}Y_{ij}=k_i\right)\\&=\prod_{i=1}^{n}\left[\frac{P(Y_{i1}=k_i, Y_{ij}=c, j=2,\ldots,M+1|X_{ij})}{\sum_{l=1}^{M+1}P(Y_{il}=k_i, Y_{ij}=c, j\neq l)}\right]\\&=\prod_{i=1}^{n}\frac{P(Y_{i1}=k_i|X_{i1})\prod_{j=2}^{M+1}P(Y_{ij}=c|X_{ij})}{\sum_{l=1}^{M+1}P(Y_{il}=k_i|X_{il})\prod_{j\neq l}P(Y_{ij}=c|X_{ij})}\end{aligned}$$

After plugging in the probability expressions (7) and (8) in the above likelihood, the $\gamma_{ik_i}^{*}$ cancel in the numerator and the denominator and the conditional likelihood simplifies to

$$L_{\text{CADJ}}=\prod_{i=1}^{n}\frac{\exp(\beta(c-k_i)X_{i1})}{\sum_{j=1}^{M+1}\exp(\beta(c-k_i)X_{ij})} \tag{9}$$

One can easily note that because of the multinomial logit structure in the transformed covariates, the prospective–retrospective equivalence for estimating β holds with this model. Also, note that for 1:M matched data, this model can be implemented easily by simply multiplying all the covariates in stratum i by $(c-k_i)$ and then using CLR in any standard software with the transformed covariates. We will call this method CADJ for future references.

Remark 4

As mentioned before, one could let β vary with the adjacent categories. In that case, one has to substitute $(c-k_i)\beta$ by $\sum_{r=1}^{c-k_i}\beta_r$ in the conditional likelihood in (9). Fitting that model is also easy by running separate CLR for the matched sets with case status k_i with transformed covariates. However, though this model has more flexibility, as a referee pointed out, this varying slope model does not have as much power advantage as the model with constant β.

4. POLYCHOTOMOUS CONDITIONAL LOGISTIC REGRESSION

We also conducted two traditional analyses without using ordinality in disease states, using a PCLR (by treating the disease categories as nominal) and usual CLR (collapsing the two subgroups within the cases as a single one). Following our previous notation, we consider c nominal levels of the disease variable, with $Y_{ij}=k$ denoting disease state k, $k=1,\ldots,c-1$ and $Y_{ij}=c$ denoting the control group. The disease probabilities for the $c-1$ case categories are modelled through $c-1$ logits as in a multinomial logistic regression model,

$$\log \frac{P[Y_{ij}=k|X_{ij}]}{P[Y_{ij}=c|X_{ij}]}=\gamma_{ik}+\beta_k X_{ij} \quad \text{for } k=1,\ldots,c-1 \tag{10}$$

For our example, $c=3$, and $\exp(\beta_1)$ signifies the odds of having *very*-low-birthweight baby for a mother who smokes relative to one who does not smoke and similarly $\exp(\beta_2)$ is the odds of having a low-birthweight child for a mother who smokes relative to one who does not. The stratum-specific nuisance parameter for disease category k in matched set i is denoted by γ_{ik}. As shown in Section 3, the conditional likelihood can be expressed as

$$L_c=\prod_{i=1}^{n}\frac{\exp(\beta_{k_i}X_{i1})}{\sum_{l=1}^{M+1}\exp(\beta_{k_i}X_{il})} \tag{11}$$

One can now maximize the above likelihood in terms of β_k, $k=1,\ldots,c-1$. Note that this is equivalent to using separate CLR for each disease-sub group. We will denote this method by PCLR. We will also implement the usual binary CLR, collapsing the two case-subgroups into a single one, which is a special case of PCLR with $c=2$.

Remark 5

The log OR estimates from these nominal models are not directly comparable with those obtained from the proportional odds model or the constant slope adjacent category logit model as all of them have different interpretations. The leverage of most ordinal models in terms of power over a nominal model is because ordinal models typically use a single parameter for exposure effect, whereas the nominal methods use category-specific parameters and lose efficiency. The gain in power with ordinal models is more when we have a moderate number of categories c when compared to nominal models or just binary collapsing [29]. Liu and Agresti [30] present a discussion in terms of gain in power for using nominal *versus* ordinal models as well as binary *versus* ordinal model. We will also observe this phenomenon in our subsequent data analysis and simulation results.

5. ANALYSIS OF REAL DATA: THE LOW-BIRTHWEIGHT STUDY

In this section, we consider a matched case–control data set coming from a low-birthweight study conducted by the Baystate Medical Center in Springfield, Massachusetts. The data set is discussed in Hosmer and Lemeshow [31, Section 1.6.2] and is used as an illustrative example of analysing a matched case–control study in Chapter 7 of their book. Low birthweight, defined as birthweight less than 2500 g, is a cause of concern for a newborn as infant mortality and birth defect rates are very high for low-birthweight babies. The data were matched according to the age of the mother. A woman's behaviour during pregnancy (smoking habits, diet, prenatal care) can greatly alter the chances of carrying the baby to full term. The goal of the study was to determine whether these variables were 'risk factors' in the clinical population served by Baystate Medical Center. Using the actual birth weight observations, we divided the cases, namely, the low-birthweight babies into two categories, *very* low (weighing less than 2000 g) and low (weighing between 2000 and 2500 g) and tried to assess the impact of smoking habits of mother on the chance of falling in the two low-birthweight categories. We consider two data sets based on this low-birthweight study with

DOI: 10.1002/sim

Table I. Analysis of low-birthweight study, one is 1:1 matched data, and the other is 1:3 matched data.

			β	
			1:1 Matched	1:3 Matched
Proportional odds model-based method	MH	Estimate	0.97	0.93
		SE	0.44	0.46
		CI	(0.11, 1.83)	(0.02, 1.83)
	PBV	Estimate	1.18	0.81
		PSD	0.44	0.40
		HPD	(0.21, 1.85)	(−0.02, 1.61)
	Unmatched analysis (single γ)	Estimate	0.82	0.89
		SE	0.37	0.43
		CI	(0.10, 1.54)	(0.05, 1.75)
	Unconditional MLE	Estimate	1.99	1.27
		SE	0.57	0.52
		CI	(0.88, 3.11)	(0.26, 2.27)
Adjacent category logit-based method	CADJ	Estimate	0.66	0.57
		SE	0.30	0.29
		CI	(0.07, 1.25)	(0.01, 1.13)
	Unmatched analysis (single γ)	Estimate	0.52	0.49
		SE	0.26	0.29
		CI	(−0.002, 1.03)	(−0.08, 1.07)
	Unconditional MLE	Estimate	1.37	0.77
		SE	0.44	0.36
		CI	(0.52, 2.23)	(0.07, 1.48)
Nominal and binary methods	PCLR	Estimate of β_1	0.85	0.63
		SE	0.69	0.71
		CI	(−0.51, 2.20)	(−0.76, 2.02)
		Estimate of β_2	1.10	1.16
		SE	0.52	0.57
		CI	(0.09, 2.11)	(0.05, 2.27)
	CLR	Estimate	1.01	0.96
		SE	0.41	0.44
		CI	(0.20, 1.82)	(0.10, 1.82)

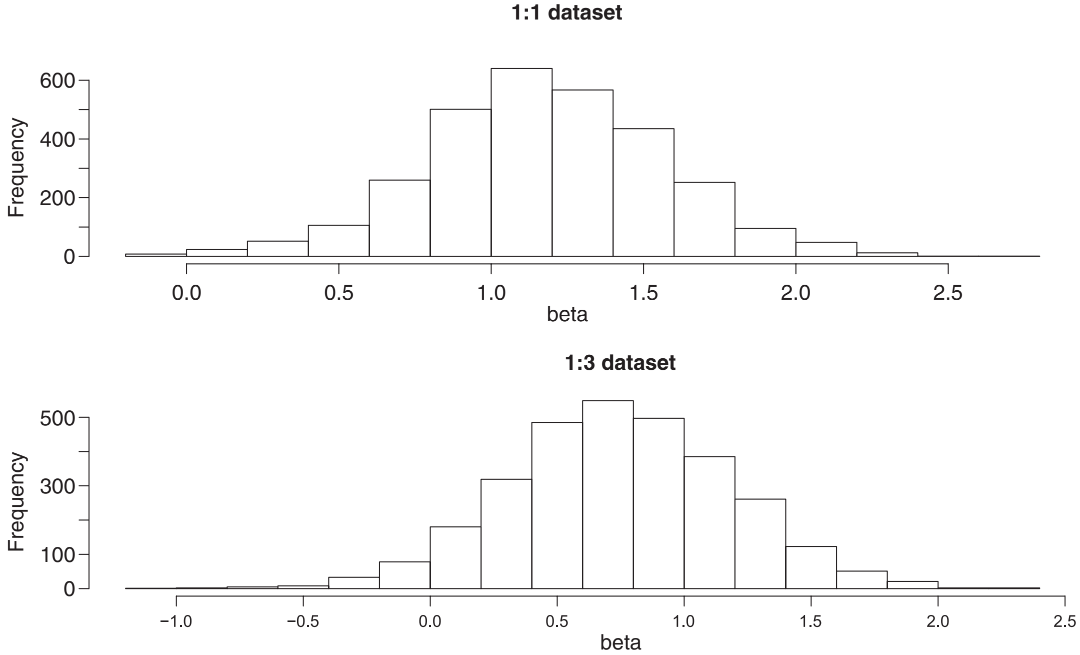
Note: SE stands for standard error, and for the Bayesian method PBV, PSD stands for standard deviation of the posterior distribution of the parameter of interest. HPD and CI denote the 95 per cent highest posterior density credible interval and confidence interval, respectively.

two different matching ratios: a 1:1 matched data set with 56 matched sets and a 1:3 matched data set with 29 matched sets. Analysis of the 1:1 and 1:3 data sets are presented in Table I.

5.1. Methods with proportional odds model

We first consider the MH method with the stratified proportional odds model in (1) as the underlying prospective model. For the 1:1 matched case–control study $\log(\hat{\theta}) = 0.97$, implying that the estimated odds of having a baby weighing below any fixed level for a smoker mother are $\exp(0.97) = 2.64$ times higher than the odds for a non-smoker mother. In the 1:3 matched case–control study $\log(\hat{\theta}) = 0.93$ (OR $= 2.53$).

DOI: 10.1002/sim


1:1 dataset

1:3 dataset

Figure 1. Posterior distribution for β under PBV method for low-birthweight study.

For the PBV method, note that the results reflect certain numerical differences when compared to the MH counterparts. In terms of prior selection for the Bayesian PBV method, for the presented results we set $\alpha_1 \equiv 0$, and use a Gamma(3.2, 0.3) prior on α_2. We choose a diffused N(0, 4) prior on β and iid N(−2, 4) prior on γ_i. We noted in other simulations that the PBV method is quite sensitive to choice of priors on γ_i and α_2, especially for the 1:1 matched data with 56 stratum-specific parameters. Figure 1 presents a plot of approximate posterior distributions of β which appears to be fairly symmetric.

If one ignored the effect of matching and used a constant stratum-specific intercept model with $\gamma_i = \gamma_0$, it is easy to obtain the MLE's of the parameters. Table I reconfirms the prior observation [2] that unmatched analysis of matched data produces conservative estimates of the relative risk parameters.

One may also be curious to know how the unconditional MLE of β performs in this case, though it is known to be inconsistent. The unconditional MLE of β is 1.992 (OR = 7.331) for the 1:1 matched study, whereas it is 1.265 (OR = 3.54) for the 1:3 data set. This again illustrates that the unconditional MLE of β is biased upwards. For $c = 2$, for 1: M matched data set it is known that the inflation factor is approximately $(M+1)/M$. We are essentially noting the same phenomenon here with the ordinal model. The bias in the unconditional ML estimates decreases as M increases.

5.2. Adjacent category logit model

With a constant β model across all pairs (which has a different interpretation than the β in (1)), the estimate of adjacent category logit β for the 1:1 data set using the CADJ method is 0.663

DOI: 10.1002/sim

(SE = 0.3). For the 1:3 data set, the CADJ method gives an estimate of β as 0.573 (SE = 0.285), making the effect of smoking marginally significant. The unconditional MLE and the estimate obtained by unmatched analysis ignoring matched-set-specific parameters show similar patterns as in the cumulative logit model.

5.3. Nominal models

The PCLR estimates indicate that smoking of mother is a more significant risk factor for having a low-birthweight child (disease group 2), but not so for having a very-low-birthweight child (disease group 1). The CLR estimates indicate that effect of smoking is more pronounced in 1:1 data set and marginally significant in 1:3 data set.

Since we do not know the truth in a real data set, we undertake a simulation study to examine the performance of the five methods: MH, PBV, CADJ, PCLR and CLR.

6. SIMULATION STUDY

We conduct two different sets of simulations. One with 1:1 matched setting and the other with 1:3 matched setting, each with 56 matched sets. We consider two disease subcategories, $Y = 1$ and 2 and, one control group $Y = 3$. The disease incidence model is assumed to follow (1) with $c = 3$. For all simulation studies we set $\alpha_2 = 1.2$ and a grid of values for $\beta = -1, -0.5, 0, 1.0, 2.0$ ($\alpha_1 = 0$ for identifiability of the models). For both 1:1 and 1:3 situations, we generate varying stratum effects $\gamma_i \overset{\text{iid}}{\sim} \mathrm{N}(-3, 4)$. Simulation results for the PBV method are presented for the same set of priors mentioned in the data analysis section.

In all simulation settings, we first generate the exposure variable X in each matched set from a Bernoulli distribution with success probability 0.41 (which was the overall prevalence of smoking mothers in the actual 1:1 matched case–control data). We then generate the disease status variable given the exposure values for the matched pairs, reflecting the matched sampling strategy under likelihood (4).

We simulate 5000 data sets under each scenario, and calculate parameter estimates and mean squared errors for MH, PBV, and CLR methods. Note that, since we are generating data from the proportional odds model in (1), it is not quite appropriate or meaningful to calculate measures like bias and MSE for CADJ or PCLR method as we do not know the true value of the parameters.

Table II contains the results. To obtain an *ad hoc* estimate of the proportion of true rejections of $H_0 : \beta = 0$, among these 5000 runs, we calculate the following proportions. For PBV method, we construct equal tailed 95 per cent credible interval of β using 2.5 per cent and 97.5 per cent sample quantiles of the corresponding posterior distribution and count the proportion of times the value zero lies outside this interval. For all other methods we first calculate the z-score by dividing the estimate with standard error and then calculate the proportion of times the absolute value of the z-score exceeds the value of 1.96.

The results indicate that PBV in general produces biased results, the bias does appear to depend on the value of β, but always persists, even in simulations with larger sample sizes (results not included here). We also found the degree of bias and precision of the Bayes estimate of β varies in terms of the prior on α_2 and γ_i. The estimated powers for PBV are also high, but at the cost of an elevated size. A part of it could be due to violation of retrospective–prospective equivalence with the prospective likelihood in (4) as discussed in Section 2.2.2. This bias is possibly

Table II. Results of the simulation study for 1:1 and 1:3 matched case–control data with varying values of β, showing mean-squared error, bias (in parentheses) and power for each method.

Value of β			MH	PBV	CLR	CADJ	PCLR(β_1)	PCLR(β_2)
-1	1:1 data	MSE(Bias)	0.25(-0.08)	0.44(-0.25)	0.44(-0.07)			
		Power	0.64	0.63	0.66	0.66	0.34	0.39
	1:3 data	MSE(Bias)	0.17(-0.07)	0.40(-0.12)	0.36(-0.01)			
		Power	0.81	0.68	0.69	0.73	0.41	0.46
-0.5	1:1 data	MSE(Bias)	0.20(-0.06)	0.36(0.05)	0.32(0.00)			
		Power	0.23	0.20	0.23	0.28	0.12	0.14
	1:3 data	MSE(Bias)	0.13(-0.04)	0.33(-0.05)	0.33(-0.01)			
		Power	0.30	0.25	0.29	0.34	0.15	0.19
0	1:1 data	MSE(Bias)	0.17(0.00)	0.36(0.24)	0.32(0.00)			
		Power/size	0.05	0.15	0.05	0.05	0.03	0.03
	1:3 data	MSE(Bias)	0.12(-0.01)	0.34(0.15)	0.27(0.00)			
		Power/size	0.05	0.09	0.05	0.05	0.04	0.04
1	1:1 data	MSE(Bias)	0.26(0.10)	0.47(0.52)	0.37(0.01)			
		Power	0.67	0.91	0.67	0.65	0.32	0.36
	1:3 data	MSE(Bias)	0.13(0.06)	0.41(0.25)	0.33(0.02)			
		Power	0.88	0.93	0.82	0.87	0.64	0.55
2	1:1 data	MSE(Bias)	0.47(0.19)	0.29(0.38)	0.50(0.05)			
		Power	0.99	1.00	1.00	0.98	0.91	0.89
	1:3 data	MSE(Bias)	0.26(0.18)	0.22(0.30)	0.42(0.04)			
		Power	1.00	1.00	1.00	1.00	0.95	0.94

Note: Results are based on 5000 runs each with 56 matched sets.

also because of the presence of so many nuisance parameters, introducing numerical instability in the Bayes estimates. The nuisance parameters are theoretically identifiable, but barely any information is available on them through the data. The MH method performs reasonably in terms of MSE and power and has the right size. Among the two methods based on the proportional odds model, MH appears to be a less subjective choice though is limited to only binary/categorical exposure.

For the CADJ method, we do not know the true value of the adjacent category log OR so we refrain from providing parameter MSE and bias. However, the method appears to have the right size and reasonable power. Similar differences with interpretation of the parameters hold for the PCLR method, but we can still note from the power estimates that it is less powerful in detecting association under the given simulation model than the other methods. The CLR estimates lose the information and interpretation of multiple disease categories. If the cumulative logit model is true and we use CLR with collapsing the disease states, we might lose some power, but the parameter estimate of β should be similar to the MH method. The loss of information from collapsing to a binary response has been discussed by Whitehead [32]. From the simulation results, it seems that compared to the MH estimate, there is not much loss of power by ignoring the ordering for the 1:1 case, however, there does appear to be some loss of power in the 1:3 case.

DOI: 10.1002/sim

7. COMMENTS

In this article we address the problem of analysing matched case–control data with ordered disease states. With advances in modern medicine and clinical research, diseases are being progressively categorized into finer subclasses, often having an intrinsic natural order. To utilize the ordinal nature of disease states, we consider using the proportional-odds model and the adjacent category logit model for stratified ordinal response. In contrast to the classical principle of using conditional likelihood to eliminate matched set-specific nuisance parameters, there is no reduction due to sufficiency in the proportional-odds model; consequently one has to deal with the nuisance parameters. We investigate two methods to tackle this issue: a Mantel–Haenszel approach (MH) and a full Bayesian approach (PBV). Our simulation study indicates that for a matched case–control study, the Bayesian method allowing for varying stratum-specific parameters in general produces biased estimates and is prior sensitive. As discussed in Section 2.2.2, the prospective likelihood based on the proportional-odds model does not acknowledge the retrospective sampling scheme. This phenomenon contributes to the bias in PBV; however, we believe that the choice of the prior distribution and sparsity of the data also contribute to this bias and sensitivity. The MH method stands superior to other methods under the proportional odds structure as it does not make direct use of the likelihood in (4). Like the ordinary MH estimator for several 2×2 tables, the Liu–Agresti MH estimator (2) is also produced from an unbiased estimating function, which could explain this difference in behaviour. The connection between the Liu–Agresti and Agresti–Lang estimator is an interesting observation in itself.

The conditional adjacent category logit model has certain attractive features as it (i) allows for prospective–retrospective equivalence due to the multiplicative intercept and odds structure in the transformed covariates, (ii) allows elimination of nuisance parameters *via* classical conditional likelihood, (iii) has more power in detecting disease–exposure association by accounting for the ordering of the disease states, (iv) can be implemented by transforming the covariates and running CLR in any software and (v) can accommodate any number of covariates, continuous or categorical because of its regression structure. However, one may want to relax the constant slope assumption for case–control studies and consequently lose some of its power advantages.

There are many interesting open questions related to this problem, for example, to theoretically investigate efficiency of the estimate formed by simultaneously fitting CLR, with binary collapsing of the ordinal response as proposed in Agresti and Lang [12] in the general regression setting of a 1:M matched data set with multiple exposures. In a preliminary study, we noted that this method of amalgamating CLR estimates from several binary collapsing of the disease states and then correcting the variance estimate works fairly well for a general regression model under 1:M matching, further research is under progress on this issue. To impose a non-parametric mixture model for the stratum effects and characterize sensible prior distributions, if at all possible, is of interest as well. Retrospective modelling as discussed in Section 2.2.2 deserves a full-length investigation. Our goal has been more modest. We simply explore issues with readily implementable and understandable methods which are of greater appeal to practitioners. R, SAS and Fortran codes for implementing the different methods are available at http://www.sph.umich.edu/bhramar/public_html/research.

ACKNOWLEDGEMENTS

The authors would like to thank Alan Agresti for a careful reading of a previous version of this manuscript. The authors are indebted to the associate editor and the two referees for their insightful comments leading

DOI: 10.1002/sim

to substantive improvement in the content and presentation of this paper. The research of Bhramar Mukherjee was partially supported by NSA young investigator grant H98230-06-1-0033.

REFERENCES

1. Agresti A. *Categorical Data Analysis* (2nd edn). Wiley: New York, 2002.
2. Breslow NE. Statistics in epidemiology: the case–control study. *Journal of the American Statistical Association* 1996; **91**:13–28.
3. Breslow NE, Day NE, Halvorsen KT, Prentice RL, Sabai C. Estimation of multiple relative risk functions in matched case–control studies. *American Journal of Epidemiology* 1978; **108**:299–307.
4. Godambe VP. Conditional likelihood and unconditional optimum estimating equations. *Biometrika* 1976; **63**: 277–284.
5. Liang K-Y, Stewart W. Polychotomous logistic regression methods for matched case–control studies with multiple case or control groups. *American Journal of Epidemiology* 1987; **125**:720–730.
6. Becher J, Jöckel KH. Bias adjustment with polychotomous logistic regression in matched case–control studies with two control groups. *Biometrical Journal* 1990; **7**:801–816.
7. Becher H. Alternative parameterization of polychotomous models: theory and application to matched case–control studies. *Statistics in Medicine* 1991; **10**:375–382.
8. Thomas DC, Goldberg M, Dewar R, Stemiatycki J. Statistical methods relating several exposure factors to several diseases in case-heterogeneity studies. *Statistics in Medicine* 1986; **5**:49–60.
9. Durbin N, Pasternack BS. Risk assessment for case–control subgroups by polychotomous logistic regression. *American Journal of Epidemiology* 1986; **6**:1101–1117.
10. Sinha S, Mukherjee B, Ghosh M. Bayesian semiparametric modeling for matched case–control studies with multiple disease states. *Biometrics* 2004; **60**:41–49.
11. Mccullagh P. A logistic model for paired comparisons with ordered categorical data. *Biometrika* 1977; **64**:449–453.
12. Agresti A, Lang JB. A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika* 1993; **80**:527–534.
13. Liu I-M, Agresti A. Mantel–Haenszel-type inference for cumulative odds ratios. *Biometrics* 1996; **52**:1222–1234.
14. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics* 1994; **50**:933–944.
15. McCullagh P. Regression models for ordinal data. *Journal of the Royal Statistical Society*, *Series B* 1980; **42**:109–142.
16. Neyman J, Scott EL. Consistent estimates based on partially consistent observations. *Econometrika* 1948; **16**:1–22.
17. Andersen EB. *Discrete Statistical Methods with Social Science Applications*. North-Holland: New York, 1980.
18. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947; **12**:153–157.
19. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 1959; **22**:719–748.
20. Mantel N. Chi-square tests with one degree of freedom: extensions of the Mantel–Haenszel procedure. *Journal of the American Statistical Association* 1963; **58**:690–700.
21. Breslow NE, Day NE. *Statistical Methods in Cancer Research*: *Volume 1*: *The Analysis of Case–Control Studies*. International Agency for Research in Cancer: Lyon, 1980.
22. Smith AFM, Roberts GO. Bayesian computation via the Gibbs Sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*, *Series B* 1993; **55**:3–23.
23. Seaman SR, Richardson S. Equivalence of prospective and retrospective models in the Bayesian analysis of case–control studies. *Biometrika* 2004; **91**:15–25.
24. Rice KM. Equivalence between conditional and mixture approaches to the Rasch model and matched case–control studies, with applications. *Journal of the American Statistical Association* 2004; **99**:510–522.
25. Prentice RL, Pyke R. Logistic disease incidence models and case–control studies. *Biometrika* 1979; **66**:403–411.
26. Neuhaus JM. Bias due to ignoring the sample design in case–control studies. *Australian and New Zealand Journal of Statistics* 2002; **44**:285–293.
27. Scott AJ, Wild CJ. Fitting regression models to case–control data by maximum likelihood. *Biometrika* 1997; **84**:57–71.
28. Kagan A. A note on the logistic link function. *Biometrika* 2001; **88**:599–601.

DOI: 10.1002/sim

29. Hu FB, Goldberg J, Hederker D, Henderson WG. Modelling ordinal responses from co-twin control studies. *Statistics in Medicine* 1998; **17**:957–970.
30. Liu I, Agresti A. The analysis of ordered categorical data: an overview and a survey of recent developments. *Test* 2005; **14**:1–73.
31. Hosmer DW, Lemeshow S. *Applied Logistic Regression* (2nd edn). Wiley: New York, 2000.
32. Whitehead J. Sample size calculations for ordered categorical data. *Statistics in Medicine* 1993; **12**:2257–2271.

DOI: 10.1002/sim