On the equivalence of posterior inference based on retrospective and prospective likelihoods: application to a case-control study of colorectal cancer

M. Ghosh,[a] J. Song,[b*†] J. J. Forster,[c] R. Mitra[c] and B. Mukherjee[d]

The paper develops a class of priors that leads to equivalent posterior inference for odds ratio parameters based on prospective and retrospective models for categorical response data. The results are applicable to both unmatched and matched case-control studies. The results hold for a general class of link functions for categorical response. The proposed method can accommodate multiple and possibly ordered disease states. The results are applied to the analysis of discrete subtypes in an ongoing case-control study of colorectal cancer. A simulation study illustrates the need for carefully considering prior choices in Bayesian analysis of data collected under retrospective design. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: case-control; colorectal cancer; multiplicative intercept; stereotype regression; stratification

1. Introduction

The main objective of case-control studies is to measure the degree of association between a certain disease (e.g., cancer) and one or more exposure variables under consideration (e.g., smoking, family history, obesity, etc.). Statistical analysis of case-control data is primarily based on the exposure-disease, or equivalently the disease-exposure, odds ratio [1] that marks the degree of association between the disease and the exposure. With advances in modern medicine and clinical diagnosis, more precise characterization of disease subtypes are often possible with the corresponding need for a more generalized odds ratio function, or equivalently more general models for the probability of disease given exposure than logistic regression models, to capture the risk heterogeneity across different subtypes.

Case-control studies are primarily retrospective in nature because one traces the exposure history of an individual or a group of individuals conditional on their outcome categories. As such, the usual prospective analysis suitable for cohort data is not obviously applicable in the case-control context. Thus, it is important to know whether a prospective analysis of the disease-exposure odds ratio parameters produces equivalent answers to the corresponding retrospective analysis.

With binary outcomes and the prospective logistic regression model, Anderson [2] provided an equivalence result for the inference of the odds ratio parameter, restricted only to discrete exposures. Continuing with this model, the classic paper of Prentice and Pyke [3] provided a more general result, where the exposure variables could be discrete, continuous, or a combination. In particular, Prentice and Pyke proved that the maximum likelihood estimator (MLE) for the odds ratio parameter based on a prospective model is equivalent to that based on a retrospective model. Moreover, this MLE is consistent even

[a]Department of Statistics, University of Florida, Gainesville, FL, U.S.A.
[b] Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, U.S.A
[c]Southampton Statistical Sciences Research Institute, University of Southampton, U.K.
[d]Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.
*Correspondence to: J. Song, Department of Statistics, University of Florida, Gainesville, FL, U.S.A.
†E-mail: song.jihyun@gmail.com

under the retrospective model, and the observed Fisher information matrix for the odds ratio parameters based on the prospective model provides valid standard error estimates under the retrospective model. The work of Prentice and Pyke [3] spurred further research in this general area. We may refer to [4–12].

Seaman and Richardson [13] first studied this equivalence of *posteriors* for the odds ratio parameter in a Bayesian context. They presented a class of priors for which the posterior for odds ratio parameters based on the prospective likelihood is equivalent to that based on the retrospective likelihood for unmatched case-control studies under the logistic regression model. Ghosh *et al.* [14] extended their work to matched case-control studies using the same class of priors as that of Seaman and Richardson [13]. Staicu [15] presented a general class of priors for the posterior equivalence of odds ratio parameters that includes the class of priors of Seaman and Richardson [13]. She also showed that adopting the priors of Seaman and Richardson [13] is necessary and sufficient for the odds ratio parameter's posterior equivalence when the nuisance parameter in the prospective likelihood, the nuisance parameter in the retrospective likelihood, and the odds ratio parameter are mutually independent. However, all these papers were restricted to the binary logistic model and discrete exposures.

The purpose of this paper is to extend the aforementioned results in two directions. First, we accommodate multiple, possibly ordered, disease states. Second, we relax the restriction to logistic regression models, by including those involving the probit and skew-symmetric links. Another interesting case that we can handle is the proportional odds model and, in particular, the cumulative logit model [16, p.322]. However, the results are still limited to models with discrete exposure variables.

A special and important case of the general model structure we consider is the multiplicative intercept model considered in [17]. Anderson [18] first introduced a further special case stereotype model and Greenland [19], Kuss [20], and Ahn *et al.* [21, 22], among others, subsequently studied it. A stereotype model lies in-between the polytomous logistic model and the proportional odds model but can represent the ordinal characteristics of a response variable under suitable order constraints on the parameters.

To establish equivalence of *posteriors* for the odds ratio parameters, starting from either the prospective or the retrospective likelihood, is important for Bayesian analysis of data collected under retrospective sampling. Direct use of the retrospective likelihood usually involves more nuisance parameters than the corresponding prospective likelihood, especially as the dimension of the exposure/covariate vector grows. For example, for binary disease states and categorical exposures, the number of nuisance parameters corresponding to the retrospective likelihood is the number of possible exposure configurations less one, whereas the number of nuisance parameters corresponding to the prospective likelihood is simply one. Once the posterior equivalence of odds ratios from the prospective and retrospective likelihoods is established, one can validly use prospective likelihood for faster computation and convergence of the numerical integration scheme needed for a full Bayes analysis, with much fewer parameters than using the retrospective likelihood. The prospective likelihood can easily be handled by standard software like WinBugs, whereas directly using the retrospective likelihood will require writing independent codes. Our characterization of priors illustrates that arbitrary priors on the nuisance parameters in the retrospective likelihood do not yield posterior equivalence. We present simulation results that indicate that with arbitrary prior choices on the nuisance parameters in the retrospective likelihood, one can end up with estimates with poor frequentist properties compared with the estimates with a prior for which the posterior equivalence holds (Table III). For a careful Bayesian analysis with categorical outcome and exposure, our results provide useful information to guide prior choices and justify the prospective likelihood for basis of Bayesian inference of retrospective data.

With improvements in clinical diagnosis and classification, different cancers are now often accompanied with detailed tumor subtype information. Moreover, sampling of cases is also being carried out according to subtype groups, and the need for models that incorporate categorical and ordered link functions under retrospective sampling is indeed important. The extension in this paper of the existing body of results to a general link function that can handle multiple ordered outcome is timely. In our example from the Molecular Epidemiology of Colorectal Cancer (MECC) study, multiple cancer stages constitute the disease subtype variable, and Bayesian methods are employed under an uncommon class of regression models called the stereotype regression models. Use of Bayesian methods appears to be most natural for this class of models, but under case-control sampling, Bayesian inference needs justification. The present paper exactly fills in that gap, provides justification to carry Bayesian analysis with a prospective likelihood, and also guides prior choices for conducting that analysis.

The outline of the remaining sections is as follows. Section 2 presents the general Bayesian equivalence result characterizing a class of priors. Conditions ensuring the propriety of the posteriors are also given in this section. Section 3 gives an equivalent formulation for the special multiplicative intercept

model. Section 4 extends the results of Sections 2 and 3 for stratified case-control studies. Section 5 presents an example of Bayesian analysis of a case-control study with multiple cancer stages under the stereotype model. Section 6 presents a simulation study. Section 7 contains some final remarks.

2. The Bayesian equivalence result

Our objective is to find a class of priors for which the posterior inference for the odds ratios remains the same under the prospective and retrospective likelihoods for any model. Consider the situation when there are $r + 1$ disease categories $d = 0, 1, \ldots, r$ and an exposure vector X that can take the values $x_1, \ldots, x_K$. Let

$$P_{dk} = P(D = d | X = x_k), \qquad d = 0, 1, \ldots, r, \quad k = 1, \ldots, K, \tag{1}$$

where $P(D = d | X = x_k)$ is any arbitrary model. Also, let

$$P(X = x_k | D = 0) = \theta_k, \qquad k = 1, \ldots, K, \tag{2}$$

where $\theta_K = 1 - \sum_{k=1}^{K-1} \theta_k$.

Then by using a technique of Satten and Kupper [23], it is easy to show that

$$P(X = x_k | D = d) = \frac{\theta_k P_{dk} P_{0k}^{-1}}{\sum_{l=1}^{K} \theta_l P_{dl} P_{0l}^{-1}}, \qquad k = 1, \ldots, K. \tag{3}$$

We denote by Z_{dk} the number of individuals with $D = d$ and $X = x_k$. Then from (1), the prospective likelihood is given by

$$L_P = \prod_{d=0}^{r} \prod_{k=1}^{K} P(D = d | X = x_k)^{Z_{dk}} = \prod_{d=0}^{r} \prod_{k=1}^{K} P_{dk}^{Z_{dk}}, \tag{4}$$

whereas from (2) and (3), the retrospective likelihood is given by

$$L_R = \prod_{k=1}^{K} \prod_{d=0}^{r} P(X = x_k | D = d)^{Z_{dk}} = \prod_{k=1}^{K} \left[\theta_k^{Z_{0k}} \prod_{d=1}^{r} \left\{ \frac{\theta_k P_{dk} P_{0k}^{-1}}{\sum_{l=1}^{K} \theta_l P_{dl} P_{0l}^{-1}} \right\}^{Z_{dk}} \right]. \tag{5}$$

We write $\alpha_d = \sum_{l=1}^{K} P_{dl}/P_{0l}$, $d = 1, \ldots, r$, and $\eta_{dk} = \alpha_d^{-1}(P_{dk}/P_{0k})$, $d = 1, \ldots, r$, $k = 1, \ldots, K$. Note that the η_{dk} are normalized odds and $\sum_{l=1}^{K} \eta_{dl} = 1$, $d = 1, \ldots, r$. The odds ratio $(P_{dk}/P_{0k})/(P_{dl}/P_{0l})$ is equivalently expressible as η_{dk}/η_{dl}. Hence, the η_{dk} are parameters of interest. As an example, for the multiplicative intercept model where $P_{dk} = \exp\{\rho_d + Q(x_k, \delta_d)\}/\sum_{j=0}^{r} \exp\{\rho_j + Q(x_k, \delta_j)\}$, where Q can be any arbitrary function of the exposure variable and parameters, one obtains the odds ratios $\eta_{dk}/\eta_{dl} = \exp[Q(x_k, \delta_d) - Q(x_l, \delta_d)]$ that do not depend on the intercept parameters ρ_d.

We now rewrite L_P with $\alpha_1, \ldots, \alpha_r$ and $\eta_{11}, \ldots, \eta_{rK}$ as

$$L_P = \left\{ \prod_{d=1}^{r} \prod_{k=1}^{K} (\alpha_d \eta_{dk})^{Z_{dk}} \right\} \prod_{k=1}^{K} \left\{ \left(1 + \sum_{d=1}^{r} \alpha_d \eta_{dk} \right)^{-\sum_{d=0}^{r} Z_{dk}} \right\}, \tag{6}$$

and rewrite L_R with $\theta_1, \ldots, \theta_K$ and $\eta_{11}, \ldots, \eta_{rK}$ as

$$L_R = \prod_{k=1}^{K} \left\{ \theta_k^{Z_{0k}} \prod_{d=1}^{r} \left(\frac{\theta_k \eta_{dk}}{\sum_{l=1}^{K} \theta_l \eta_{dl}} \right)^{Z_{dk}} \right\} \tag{7}$$

because $P_{dk}/P_{0k} = \alpha_d \eta_{dk}$ and $\sum_{l=1}^{K} \eta_{dl} = 1$. We also write $\alpha = (\alpha_1, \ldots, \alpha_r)$, $\theta = (\theta_1, \ldots, \theta_{K-1})$, $\eta = (\eta_{11}, \ldots, \eta_{1K}, \ldots, \eta_{r1}, \ldots, \eta_{rK})$. Recall that $\theta_K = 1 - \sum_{k=1}^{K-1} \theta_k$. The main theorem of this section is as follows.

Theorem 1

Consider the prior $\pi(\alpha, \eta) \propto \left(\prod_{d=1}^{r} \alpha_d^{-1}\right) \pi_0(\eta)$ for L_P and the prior $\pi(\theta, \eta) \propto \left(\prod_{k=1}^{K} \theta_k^{-1}\right) \pi_0(\eta)$ for L_R, where π_0 is an arbitrary but proper prior. Then the posterior of η is the same under L_P or L_R. Also, suppose $Z_{0k} \geqslant 1 (k = 1, \ldots, K)$, that is, every exposure category consists of at least one control. Then the posterior of η is proper.

Proof

Proof is given in Appendix A. □

Consider the situation in which one obtained his or her data from a case-control study and believes that the prior of θ and η is $\pi(\theta, \eta) \propto \left(\prod_{k=1}^{K} \theta_k^{-1}\right) \pi_0(\eta)$ where π_0 is any distribution. The standard retrospective Bayesian analysis corresponding to case-control studies will use the retrospective likelihood L_R and the aforementioned prior $\pi(\theta, \eta)$ to obtain the posterior of η. Theorem 1 tells that using a prospective likelihood L_P and a prior $\pi(\alpha, \eta) \propto \left(\prod_{d=1}^{r} \alpha_d^{-1}\right) \pi_0(\eta)$ also provides an equivalent posterior of η to that from the standard retrospective Bayesian analysis. Noting that an odds ratio is a function of η, the posterior equivalence of odds ratios from the prospective and retrospective likelihoods holds. Also, Theorem 1 generalizes Seaman and Richardson [13] who considered binary disease states, that is, $d = 0, 1$, and the logit link describing P_{dk}.

Our characterization result for posterior equivalence yields Dirichlet$(0, \ldots, 0)$ as the prior for the retrospective nuisance parameter vector θ. If one chooses a Dirichlet$(a_1, \ldots, a_K)$ prior with $(a_1, \ldots, a_k)$ not equal to $(0, \ldots, 0)$ for θ, the posterior equivalence will not hold. As will be evident in the appendix, writing $\theta_k = \gamma_k / \sum_{j=1}^{K} \gamma_j$, the γ_k in (14) are a posteriori independent exponentials with shape parameters $a_k + \sum_{d=0}^{r} Z_{dk}$ and a common scale parameter. The exact L_P cannot be obtained by integrating out the γ_k.

3. Multiplicative intercept model

Multiplicative intercept models are quite widely used for the analysis of categorical data and, in particular, for the analysis of case-control data. For such a model,

$$P_{dk} = P(D = d | X = x_k) = \frac{\exp\{\rho_d + Q(x_k, \delta_d)\}}{\sum_{m=0}^{r} \exp\{\rho_m + Q(x_k, \delta_m)\}}. \tag{8}$$

To avoid nonidentifiability, we assume without loss of generality $\rho_0 = 0$ and $Q(x_k, \delta_0) = 0$ for all $k = 1, \ldots, K$. For this special model, $(\eta_{11}, \ldots, \eta_{rK})$ is a function of $(\delta_1, \ldots, \delta_r)$. Consider a prior $\pi(\rho_1, \ldots, \rho_r, \delta_0, \ldots, \delta_r) \propto \pi_0(\delta_0, \ldots, \delta_r)$ for the prospective likelihood in (4) and a prior $\pi(\delta_0, \ldots, \delta_r, \theta) \propto \left(\prod_{k=1}^{K} \theta_k^{-1}\right) \pi_0(\delta_0, \ldots, \delta_r)$ for the retrospective likelihood in (7) with any proper prior π_0. Then the posterior of $(\delta_1, \ldots, \delta_r)$ based on the prospective likelihood is equivalent to that based on the retrospective likelihood by Theorem 1. The proof is given in Appendix B.

A special case of (8) is the general multinomial logistic model where $Q(x_k, \delta_d) = x_k^T \delta_d$. The stereotype model [18] is another special case with $Q(x_k, \delta_d) = \phi_d x_k^T \beta$, where $0 = \phi_0 < \phi_1 < \cdots < \phi_{r-1} < \phi_r = 1$ or $1 = \phi_0 > \phi_1 > \cdots > \phi_{r-1} > \phi_r = 0$. This is one of the simplest models to analyze ordinal data including the adjacent category model [16, p.318] where $\phi_d = (r - d)/r$.

4. Posterior equivalence for stratified case-control data

In this section, we extend the results of Section 2 for stratified case-control data. The need for stratification often arises in case-control studies to eliminate any potential confounding effects. For instance, often it is necessary to stratify case-control data based on gender, age, race, and ethnicity. An example of such stratified data appears in the next section. For Bayesian analysis of matched case-control data, see [24–27].

Ghosh *et al.* [15] established this posterior equivalence for binary case-control data assuming a logistic regression model. Once again, we provide a two-fold extension of their results: first, by considering multiple and possibly ordered disease states; second, by considering an arbitrary link not restricted to the logit link.

To this end, we consider T strata, $1, \ldots, T$ and denote the stratum indicator by S. As in Section 2, we have $(r + 1)$ disease categories $d = 0, 1, \ldots, r$, and K exposure categories $x_1, \ldots, x_K$. Let Z_{sdk} denote the number of individuals with $D = d$ and $X = x_k$ in stratum s. A prospective likelihood is given by

$$L_P = \prod_{s=1}^{T} \prod_{d=0}^{r} \prod_{k=1}^{K} P_{sdk}^{Z_{sdk}}, \tag{9}$$

where $P_{sdk} = P(D = d | S = s, X = x_k)$, $s = 1, \ldots, T$, $d = 0, 1, \ldots, r$, $k = 1, \ldots, K$. A retrospective likelihood is given by

$$L_R = \left\{ \prod_{s=1}^{T} \prod_{k=1}^{K} \theta_{sk}^{Z_{s0k}} \right\} \left[\prod_{s=1}^{T} \prod_{d=1}^{r} \prod_{k=1}^{K} \left\{ \frac{\theta_{sk} P_{sdk} P_{sok}^{-1}}{\sum_{l=1}^{K} \theta_{sl} P_{sdl} P_{sol}^{-1}} \right\}^{Z_{sdk}} \right], \tag{10}$$

where $P(X = x_k | D = 0, S = s) = \theta_{sk}$, $\theta_{sK} = 1 - \sum_{k=1}^{K-1} \theta_k$, $s = 1, \ldots, T$, $k = 1, \ldots, K$.

Now consider, as in Section 2, the reparameterization $\alpha_{sd} = \sum_{l=1}^{K} P_{sdl}/P_{s0l}$ and $\eta_{sdk} = \alpha_{sd}^{-1}$ (P_{sdk}/P_{s0k}), $s = 1, \ldots, T$, $d = 1, \ldots, r$, and $k = 1, \ldots, K$. We denote $\eta = (\eta_{111}, \ldots, \eta_{TrK})$. We can note that L_P and L_R can be written with $\alpha_{11}, \ldots, \alpha_{Tr}, \eta$ and $\theta_{11}, \ldots, \theta_{TK-1}, \eta$ respectively. Consider the prior

$$\pi(\alpha_{11}, \ldots, \alpha_{Tr}, \eta) \propto \left(\prod_{s=1}^{T} \prod_{d=1}^{r} \alpha_{sd}^{-1} \right) \pi_0(\eta)$$

for L_P in (9) and the prior

$$\pi(\theta_{11}, \ldots, \theta_{TK-1}, \eta) \propto \left(\prod_{s=1}^{T} \prod_{k=1}^{K} \theta_{sk}^{-1} \right) \pi_0(\eta)$$

for L_R in (10), where π_0 is a proper prior. Then it turns out that the posterior of η is the same whether it is generated from L_P or L_R, and the proof is similar to the one in Section 2.

For the multiplicative intercept model in Section 3, $P_{sdk} = \exp\{\rho_{sd} + Q(x_k, \delta_d)\} / \sum_{m=0}^{r} \exp\{\rho_{sm} + Q(x_k, \delta_m)\}$. Assume without loss of generality, $\rho_{s0} = 0$ for all $s = 1, \ldots, T$ and $Q(x_k, \delta_0) = 0$ for all $k = 1, \ldots, K$. Consider the prior $\pi(\rho_{11}, \ldots, \rho_{Tr}, \delta_0, \ldots, \delta_r) \propto \pi_0(\delta_0, \ldots, \delta_r)$ with any proper prior π_0 for the prospective likelihood and the prior $\pi(\delta_0, \ldots, \delta_r, \theta_{11}, \ldots, \theta_{TK-1}) \propto \left(\prod_{s=1}^{T} \prod_{k=1}^{K} \theta_{sk}^{-1} \right)$ $\pi_0(\delta_0, \ldots, \delta_r)$ for the retrospective likelihood. Then the posterior equivalence of $(\delta_1, \ldots, \delta_r)$ holds as in Section 3. We use this posterior equivalence of $(\delta_1, \ldots, \delta_r)$ in Section 5 to do data analysis for certain stereotype models.

5. Example: analysis of colorectal cancer data

The dataset we consider is a population-based case-control study of colorectal cancer in Israel from 1998 to 2004. The MECC Study is a rich resource of molecular, environmental, dietary, and behavioral risk factors along with basic personal and demographic information [28]. We now apply the results of the previous section to analyze a subset of the data with 1066 cases and 1337 controls with stage information recorded for the cases. We carried out cancer staging on the basis of tumor nodes and metastasis criteria. Among the cases, 665 are in stages 1 and 2 ($d = 1$) and 401 are in stages 3 and 4 ($d = 2$). Thus, the number of disease categories is 3, with $d = 0$ standing for all 1337 controls. The objective is to explore the association between colorectal cancer stage with the frequency of grilled red-meat intake and physical activity. Ethnicity may also affect risk of colorectal cancer, and we used it for stratification. We restricted our analysis to two ethnic groups, Ashkenazi and Sephardi, that constitute stratum 1 ($s = 1$) with 1799 observations and stratum 2 ($s = 2$) with 604 observations, respectively. The exposure vector consists of two variables. RM is coded as 1 if the subject eats grilled red meat at least once a week and 0 otherwise. We also use a binary variable, PA, whose value is 1 if the subject is physically active and is 0 otherwise. Thus, the four possible values of the exposure vector with these two dichotomous variables are (0,0), (1,0), (0,1), (1,1), which we denote by x_1, x_2, x_3, x_4.

Stereotype models are useful here owing to the ordered nature of disease states. We use $P_{sdk} = \exp\left(\rho_{sd} + \phi_d x_k^T \beta\right) / \sum_{m=0}^{2} \exp\left(\rho_{sm} + \phi_m x_k^T \beta\right)$ with $\beta = (\beta_1, \beta_2)^T$, $\rho_{10} = \rho_{20} = 0$, $0 = \phi_0 < \phi_1 < \phi_2 = 1$ for $s = 1, 2, d = 0, 1, 2, k = 1, 2, 3, 4$. Noting that

$$\eta_{dk}/\eta_{dk'} = \exp(\phi_d(x_k - x_{k'})^T \beta), \tag{11}$$

odds ratios are functions only of the $\phi_d (d = 0, 1, 2)$, $x_k (k = 1, \ldots, 4)$ and β. Now owing to the equivalence of the posteriors of η_{dk} derived from both the prospective and retrospective likelihoods as mentioned in the previous section and (11), we work only with the prospective model for inference about ϕ_1, β, and odds ratios: $\exp(\phi_1\beta_1), \exp(\phi_1\beta_2), \exp(\beta_1), \exp(\beta_2)$.

As discussed in Section 4, in the general context of multiplicative intercept models, we use independent uniform priors for $\rho_{11}, \rho_{12}, \rho_{21}$, and ρ_{22}. We also use several normal priors: priors 1–5 for β and a uniform $(0,1)$ prior for ϕ_1. For prior 1, 0 is used as a prior mean, and the prior covariance matrix is taken as the inverse of observed per unit information matrix multiplied by the total number of observations. In this prior covariance matrix, the variances for β_1 and β_2 are 59 and 26, respectively, and the correlation between β_1 and β_2 is -0.31. For priors 2–4, the prior mean is set to be 0, and prior variances for β_1 and β_2 are set to be 50, whereas the prior correlation between β_1 and β_2 varies from -0.4 to 0, to 0.2. For prior 5, the MLE is used as a prior mean, and the covariance matrix in prior 1 is used as a prior covariance matrix. Detailed description for priors 1–5 is given in Table I footnote. With these priors, the equivalence in Section 4 holds.

We need to mention that the MLE and covariance matrix estimate used in priors 1 and 5 are obtained without the constraint $0 = \phi_0 < \phi_1 < \phi_2 = 1$ following the method proposed by Kuss [20]. This is mainly because obtaining maximum likelihood (ML) estimates is computationally difficult with the presence of the constraint in that if ϕ_1 is close to ϕ_0 or ϕ_2, numerical methods are less likely to find the optimal solution (see Section 3 of [21]). However, the advantage of the Bayesian approach is that the constraint imposed through a prior is reflected in the posterior.

Seaman and Richardson [13] recommended using caution while using uniform priors on the intercept parameters when stratum sizes are small. They observe that such flat priors can lead to biased estimates for the odds ratio parameters if a stratum size is less than 40 subjects. In our data, stratum sizes are substantially larger ($\geqslant 604$); hence, the posterior mean of our analysis is less likely to be biased because of the use of uniform priors.

We implemented the Bayesian method with Gibbs sampling. The full conditionals are given in Appendix C. Owing to the log-concavity of the conditionals, we used adaptive rejection sampling for all parameters [29]. For each analysis, we performed a total of 50,000 iterations and discarded the initial 30,000 as burn in. Graphical analysis of the MCMC sample indicated satisfactory convergence, which was also confirmed by the method proposed by Geweke [30]. A summary of MCMC samples including posterior means and 95% HPD regions for β_1, β_2, ϕ_1, and odds ratios is given in Table I. Also, MLEs and 95% asymptotic CIs are given in Table I for the purpose of comparison.

On the basis of our analysis, RM has positive association with colorectal cancer controlling for PA and ethnicity regardless of priors for β. PA is negatively associated with colorectal cancer with any prior for β when one controls for RM and ethnicity. These results are consistent with those from the ML approach. In particular, when prior 1 is used, the posterior means for β_1 and β_2 are very close to the MLEs. Also, the endpoints of 95% HPD regions for β_1 and β_2 are very close to those of 95% asymptotic CIs. However, for ϕ_1, the posterior mean is a little different from the MLE. The 95% HPD region for ϕ_1 is included in the 95% asymptotic CI and much shorter than the latter. These results show that an analysis without the constraint $0 = \phi_0 < \phi_1 < \phi_2 = 1$ may give a much wider interval estimate for ϕ_1 than the one with this constraint even though point estimators are somewhat close to each other.

Comparing results for priors 2–4, we can assess prior sensitivity. We find that the posterior mean for β_1 changes substantially from 1.08 to 0.82, to 0.66 as the prior correlation between β_1 and β_2 varies from -0.4 to 0, to 0.2. The 95% HPD region shifts left substantially as well. However, the posterior means and the 95% HPD regions for β_2 and ϕ_1 do not change very much. Accordingly, posterior means and 95% HPD regions for odds ratios involving β_1 change substantially whereas the ones not involving β_1 do not.

Comparing results for priors 1 and 5, we find that when the prior covariance for β is a covariance matrix estimate obtained from the ML method multiplied by the total number of observations, the shift of the prior mean for β from 0 to the MLE has almost no impact on posterior means and 95% HPD regions for all parameters of interest.

Table I. Posterior means and HPD regions for β_1, β_2, ϕ_1, and odds ratios with various normal priors for β, and their MLEs and asymptotic CIs.

	Prior 1		Prior 2		Prior 3	
	Posterior mean	95% HPD region	Posterior mean	95% HPD region	Posterior mean	95% HPD region
β_1 RM	1.08	(0.78,1.37)	1.08	(0.79,1.37)	0.82	(0.54,1.16)
β_2 PA	−0.61	(−0.81,−0.40)	−0.61	(−0.84,−0.41)	−0.59	(−0.79,−0.38)
ϕ_1 Cancer stage (CS) = 1–2	0.85	(0.70,1.00)	0.86	(0.70,1.00)	0.89	(0.74,1.00)
$\exp(\phi_1\beta_1)$ Odds ratio of RM for CS = 1–2	2.53	(1.88,3.19)	2.53	(1.91,3.21)	2.09	(1.58,2.77)
$\exp(\phi_1\beta_2)$ Odds ratio of PA for CS = 1–2	0.60	(0.49,0.71)	0.60	(0.49,0.70)	0.60	(0.49,0.71)
$\exp(\beta_1)$ Odds ratio of RM for CS = 3–4	2.99	(2.16,3.88)	2.98	(2.13,3.88)	2.29	(1.69,3.15)
$\exp(\beta_2)$ Odds ratio of PA for CS = 3–4	0.54	(0.43,0.66)	0.55	(0.43,0.66)	0.56	(0.45,0.68)

	Prior 4		Prior 5		ML	
	Posterior mean	95% HPD region	Posterior mean	95% HPD region	MLE	95% CI
β_1 RM	0.66	(0.41,0.82)	1.09	(0.81,1.40)	1.06	(0.76,1.37)
β_2 PA	−0.57	(−0.76,−0.38)	−0.62	(−0.83,−0.42)	−0.61	(−0.81,−0.40)
ϕ_1 Cancer stage (CS) = 1–2	0.91	(0.77,1.00)	0.85	(0.69,1.00)	0.90	(0.65,1.14)
$\exp(\phi_1\beta_1)$ Odds ratio of RM for CS = 1–2	1.83	(1.45,2.10)	2.55	(1.91,3.24)	2.59	(1.91,3.28)
$\exp(\phi_1\beta_2)$ Odds ratio of PA for CS = 1–2	0.60	(0.50,0.71)	0.59	(0.49,0.72)	0.58	(0.47,0.69)
$\exp(\beta_1)$ Odds ratio of RM for CS = 3–4	1.95	(1.46,2.2)	3.02	(2.16,3.93)	2.90	(2.00,3.80)
$\exp(\beta_2)$ Odds ratio of PA for CS = 3–4	0.57	(0.46,0.67)	0.55	(0.43,0.65)	0.55	(0.43,0.66)

$$N\left(0,\begin{bmatrix}59 & -12\\-12 & 26\end{bmatrix}\right),\ N\left(0,\begin{bmatrix}50 & -20\\-20 & 50\end{bmatrix}\right),\ N\left(0,\begin{bmatrix}50 & 0\\0 & 50\end{bmatrix}\right),\ N\left(0,\begin{bmatrix}50 & 10\\10 & 50\end{bmatrix}\right),\text{ and }N\left(\begin{bmatrix}1.06\\-0.61\end{bmatrix},\begin{bmatrix}59 & -12\\-12 & 26\end{bmatrix}\right)$$ are used as priors of β for priors 1–5, respectively.

HPD, highest posterior density; MLE, maximum likelihood estimator.

Setting the prior for β as prior 5, we also did data analysis with beta(2,5), beta(5,5), and beta(5,2) priors for ϕ_1. Table II shows that the posterior means and 95% HPD intervals for β_1 and β_2 are similar to those under a uniform prior for ϕ_1. However, the posteriors for the category specific scores, namely ϕ_1, somewhat vary.

6. Simulation studies

Looking at the unstratified retrospective likelihood in (7), we can note that assigning a Dirichlet $(a_1, \ldots, a_K)$ prior on θ plays a role of adding $a_k - 1$ controls to the kth exposure category in the likelihood. In this sense, using a Dirichlet$(0, \ldots, 0)$ prior for θ removes one control from each exposure category. Also, using an arbitrary proper Dirichlet prior for θ adds arbitrary numbers of controls to each exposure category. This can possibly lead to poor performance of the Bayes estimator of odds ratio parameters as a posterior mean causing a large bias and mean squared error (MSE), whereas the Bayes estimator with Dirichlet$(0, \ldots, 0)$ prior for θ does not. Considering that with a Dirichlet$(0, \ldots, 0)$ prior for θ, the equivalence of posteriors from the prospective and retrospective likelihoods holds for the odds ratio parameters by Theorem 1, it follows that prospective Bayes estimates from the prospective likelihood can perform better than the retrospective Bayes estimates from the retrospective likelihood when the prior for θ is an arbitrary proper Dirichlet prior. In a similar way, a retrospective Bayes estimator with an arbitrary proper Dirichlet prior for the nuisance parameter vectors $\theta_s = (\theta_{s1}, \ldots, \theta_{sK}), s = 1, \ldots, T$ in the stratified retrospective likelihood in (10), can be seriously biased and inefficient whereas the prospective Bayes estimator from the prospective likelihood is not. In this section, we illustrate this point by simulation studies.

We simulated case-control studies based on the estimated stratified stereotype model in the Section 5. We used the posterior means with prior 1 in Table I for the parameters in the prospective model. We chose arbitrarily the marginal probability functions of the exposure X in each stratum. They were $\Pr(X = x_k | S = s) = 0.25$ for $k = 1, \ldots, 4$ and $s = 1, 2$. We obtained a retrospective model using Bayes rule. We generated 100 sets of data from the retrospective model. We used a number of observations in each disease state and each stratum in the MECC data for the sample size.

We performed a prospective Bayesian analysis with each simulated data set. For the prior, we used $\pi_1(\beta, \phi_1, \rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}) \propto \pi_{01}(\beta, \phi_1)$ where π_{01} is a product of a bivariate normal with mean vector 0 and covariance matrix $\begin{pmatrix} 59 & -12 \\ -12 & 26 \end{pmatrix}$ for β and Uniform$(0, 1)$ for ϕ_1. This is prior 1 in Table I. Note that the posterior from this prospective Bayesian analysis with π_1 is equivalent to that from the retrospective likelihood with the prior $\pi(\beta, \phi_1, \theta_1, \theta_2) \propto \left(\prod_{s=1}^2 \prod_{k=1}^4 \theta_{sk}^{-1} \right) \pi_{01}(\beta, \phi_1)$ by the equivalence result in Section 4. We performed two retrospective Bayesian analyses with proper Dirichlet priors for θ_1 and θ_2 as well. For one retrospective Bayesian analysis, we used a prior $\pi_2(\beta, \phi_1, \theta_1, \theta_2) \propto \left(\prod_{s=1}^2 \prod_{k=1}^4 \theta_{sk}^{a_{sk}-1} \right) \pi_{01}(\beta, \phi_1)$ with $(a_{s1}, \ldots, a_{s4}) = (20, 40, 20, 40)$ for $s = 1, 2$. The other retrospective Bayesian analysis used $\pi_3(\beta, \phi_1, \theta_1, \theta_2) \propto \left(\prod_{s=1}^2 \prod_{k=1}^4 \theta_{sk}^{a_{sk}-1} \right)$

Table II. Sensitivity analysis under prior choice for ϕ_1: posterior means and HPD regions for β_1, β_2, ϕ_1 with beta priors for ϕ_1 and prior 5 in Table I for β.

	Prior for ϕ_1	Posterior mean	95% HPD region
β_1	beta(5,2)	1.10	(0.81, 1.40)
	beta(5,5)	1.14	(0.84, 1.43)
	beta(2,5)	1.15	(0.86, 1.45)
β_2	beta(5,2)	−0.62	(−0.83, −0.41)
	beta(5,5)	−0.65	(−0.86, −0.44)
	beta(2,5)	−0.66	(−0.88, −0.44)
ϕ_1	beta(5,2)	0.83	(0.69, 0.98)
	beta(5,5)	0.75	(0.60, 0.89)
	beta(2,5)	0.72	(0.57, 0.87)

Table III. Posterior mean, bias, and MSE from simulation studies for β_1, β_2, *and* ϕ_1 with the priors π_1, π_2, and π_3.

	true value		prospective Bayesian analysis with π_1	retrospective Bayesian analysis with π_2	retrospective Bayesian analysis with π_3
β_1	1.08	Posterior mean	1.10	0.92	1.02
		Estimated bias	0.02	−0.16	−0.06
		Estimated MSE	0.010	0.035	0.012
β_2	−0.61	Posterior mean	−0.62	−0.57	−0.46
		Estimated bias	−0.01	0.04	0.15
		Estimated MSE	0.011	0.012	0.032
ϕ_1	0.85	Posteruir mean	0.83	0.80	0.81
		Estimated bias	−0.02	−0.05	−0.04
		Estimated MSE	0.006	0.009	0.008

$\pi_{01}(\beta, \phi_1)$ with $(a_{s1}, \ldots, a_{s4}) = (40, 40, 20, 20)$ for $s = 1, 2$ for the prior. We give the estimated bias and MSE of Bayes estimators for β and ϕ_1 with these three Bayesian analyses in Table III. The prospective Bayes estimator with π_1 has very small estimated biases, that is, 0.02, −0.01, and −0.02 for β_1, β_2, and ϕ_1, respectively, and has the smallest estimated bias and MSE among the three Bayes estimators for all β_1, β_2, and ϕ_1.

The estimated biases of the retrospective Bayes estimator with π_2 and π_3 for β_1 are −0.16 and −0.06, respectively. Considering that the prospective Bayes estimator with π_1 is biased only by 0.02, one can note that the retrospective Bayes estimator with π_2 is biased substantially. This result can be explained as follows. Under the prior π_2, θ_1 and θ_2 have independent Dirichlet(20,40,20,40) priors. Because the second and fourth exposure categories have value one for RM and other categories have value zero for RM, one can note that the use of the Dirichlet(20,40,20,40) prior for θ_1 and θ_2 played a role of adding 38 non-red-meat eaters and 78 red-meat eaters to the control group in each stratum. Considering in the simulation sampling scheme, the expected numbers of the non-red-meat eater and red-meat eater in the control group were 599 and 352 in stratum 1, and 232 and 154 in stratum 2, respectively; using the Dirichlet(20,40,20,40) prior for θ_1 and θ_2 resulted in the considerable increase of the ratio of red-meat eaters over non-red-meat eaters in the control group. This led to weakening the association between RM and colorectal cancer in the posterior and caused a large negative bias of the Bayes estimator for β_1. This estimator is also most inefficient among three Bayes estimators for β_1 because it has the largest estimated MSE.

The estimated bias of the retrospective Bayes estimator for β_2 with π_3 is 0.15, whereas the retrospective Bayes estimator with π_2 and the prospective Bayes estimator with π_1 are 0.04 and −0.01, respectively. Also, the retrospective Bayes estimator with π_3 has the largest estimated MSE among three Bayes estimators for β_2. Note that for π_3, we used independent Dirichlet(40,40,20,20) priors for θ_1 and θ_2. Again, because the first and second exposure categories have value zero for PA and the other categories have value one for PA, this choice of prior adds 78 physically inactive individuals and 38 physically active individuals to the control group in each stratum. Considering the expected numbers of the physically inactive and active individuals in the control group were 409 and 542 in stratum 1, and 170 and 216 in stratum 2, respectively, under the simulation sampling scheme; using π_3 for the prior for the retrospective Bayesian analysis weakened the association between PA and colorectal cancer by decreasing the ratio of physically active individuals over the physically inactive individuals in the control group. This caused large bias and MSE for the retrospective Bayes estimator of β_2 with π_3.

For ϕ_1, the estimated biases do not vary much depending on the choice of priors. Those are −0.02, −0.05, and −0.04 for π_1, π_2, and π_3, respectively. The estimated MSE is also small for all three Bayes estimators ranging from 0.006 to 0.009.

7. Summary

The main objective of this paper is to show that the posterior equivalence of the odds ratio parameters based on prospective and retrospective likelihoods holds both for unmatched and matched case-control studies for a class of priors. We do not insist that one must use this class of priors. For example, with a properly elicited prior, if available, one can do a subjective Bayesian analysis. We simply make the

point that an analyst can do either prospective or retrospective inference once he or she uses this class of priors. Also, giving a colorectal cancer study data analysis as an example, we show that our priors can be used to analyze real data. The results generalize those of Seaman and Richardson [13] and of Ghosh *et al.* [14]. We included the multiplicative intercept model as a special case. These results are important for practitioners who are planning to conduct Bayesian analysis of case-control data. With advancement in technology and diagnosis, histological and morphological characterization of case subtypes are becoming increasingly common. Studies are being carried out where sampling is conditional on subtypes. To have an equivalence result that accommodates general link functions and allows for using a simpler prospective likelihood as the basis of Bayesian inference is indeed important in current analytic context. Moreover, our simulation results in Table III show that an arbitrary selection of priors on the retrospective nuisance parameters can affect posterior inference on the parameter of interest and lead to undesirable frequentist properties of the resultant Bayes estimate. Arbitrary choice of priors does not lead to posterior equivalence from the prospective and retrospective models. Thus, the paper provides many foundational insights for a careful Bayesian analysis of data collected under general outcome-dependent sampling designs using arbitrary link functions.

One major limitation of such equivalence results is that the exposure variables are assumed to be discrete. A natural future direction is to consider the case of arbitrary exposures, discrete, continuous, or mixed. However, this result does not follow as a limiting case of the discrete exposure result and thus remains a challenging open problem.

Appendix A. Proof for Theorem 1

Following Seaman and Richardson [13] or Ghosh *et al.* [14], we begin with the augmented model $Z_{dk}|\lambda_{dk} \overset{ind}{\sim} \text{Poisson}(\lambda_{dk})$, where

$$\log(\lambda_{dk}) = \log \alpha_d + \log \gamma_k + \log \eta_{dk}, \qquad d = 1, \ldots, r, k = 1, \ldots, K,$$

in which $\gamma_k = \psi \, \theta_k (k = 1, \ldots, K), \theta_K = 1 - \sum_{k=1}^{K-1} \theta_k, \psi > 0, \sum_{k=1}^{K} \eta_{dk} = 1, \ d = 1, \ldots, r$, and

$$\log(\lambda_{0k}) = \log \gamma_k, \qquad k = 1, \ldots, K. \tag{12}$$

Then the augmented likelihood is given by

$$L_A \propto \exp\left\{-\sum_{k=1}^{K} \gamma_k \left(1 + \sum_{d=1}^{r} \alpha_d \eta_{dk}\right)\right\} \prod_{k=1}^{K} \gamma_k^{\sum_{d=0}^{r} Z_{dk}} \prod_{d=1}^{r} \alpha_d^{\sum_{k=1}^{K} Z_{dk}} \prod_{d=1}^{r} \prod_{k=1}^{K} \eta_{dk}^{Z_{dk}}. \tag{13}$$

Suppose ψ is independent of θ and has a prior $\pi(\psi) \propto \psi^{-1}$. Also, the prior for θ is $\pi(\theta) \propto \prod_{k=1}^{K} \theta_k^{-1}$. Letting $\gamma = (\gamma_1, \ldots, \gamma_K)$, the prior for γ becomes $\pi(\gamma) \propto \prod_{k=1}^{K} \gamma_k^{-1}$ because the Jacobian of transformation of $\gamma_k = \psi \theta_k \ (k = 1, \ldots, K)$ is ψ^{-K+1}. Now consider the joint prior $\pi(\alpha, \gamma, \eta) \propto \left(\prod_{d=1}^{r} \alpha_d^{-1}\right) \left(\prod_{k=1}^{K} \gamma_k^{-1}\right) \pi_0(\eta)$. The posterior based on the likelihood given in (13) and the prior $\pi(\alpha, \gamma, \eta)$ are given by

$$\pi(\alpha, \gamma, \eta | z) \propto \exp\left\{-\sum_{k=1}^{K} \gamma_k \left(1 + \sum_{d=1}^{r} \alpha_d \eta_{dk}\right)\right\} \prod_{k=1}^{K} \gamma_k^{\sum_{d=0}^{r} Z_{dk}-1} \prod_{d=1}^{r} \alpha_d^{\sum_{k=1}^{K} Z_{dk}-1}$$

$$\times \left(\prod_{d=1}^{r} \prod_{k=1}^{K} \eta_{dk}^{Z_{dk}}\right) \pi_0(\eta). \tag{14}$$

Now, integrating out γ, from (14), one obtains

$$\pi(\eta, \alpha | z) \propto \left\{ \prod_{d=1}^{r} \prod_{k=1}^{K} (\alpha_d \eta_{dk})^{Z_{dk}} \prod_{k=1}^{K} \left(1 + \sum_{d=1}^{r} \alpha_d \eta_{dk}\right)^{-\sum_{d=0}^{r} Z_{dk}} \right\} \left(\prod_{d=1}^{r} \alpha_d^{-1} \right) \pi_0(\eta)$$

$$= L_P \left(\prod_{d=1}^{r} \alpha_d^{-1} \right) \pi_0(\eta), \tag{15}$$

by (6). Next, integrating out α, it follows from (14) that

$$\pi(\gamma, \eta | Z) \propto \exp\left(-\sum_{k=1}^{K} \gamma_k\right) \prod_{k=1}^{K} \gamma_k^{\sum_{d=0}^{r} Z_{dk}} \prod_{d=1}^{r} \prod_{k=1}^{K} \eta_{dk}^{Z_{dk}} \prod_{k=1}^{K} \gamma_k^{-1}$$

$$\times \prod_{d=1}^{r} \left(\sum_{k=1}^{K} \gamma_k \eta_{dk} \right)^{-\sum_{k=1}^{K} Z_{dk}} \pi_0(\eta). \tag{16}$$

We now write ψ and θ instead of γ using $\gamma_k = \psi \theta_k$ $(k = 1, \ldots, K)$. Because the Jacobian of transformation is ψ^{K-1}, the joint posterior of (ψ, θ, η) is

$$\pi(\psi, \theta, \eta | z) \propto \exp(-\psi) \psi^{\sum_{k=1}^{K} Z_{0k} - 1} \left(\prod_{d=1}^{r} \prod_{k=1}^{K} \eta_{dk}^{Z_{dk}} \right) \prod_{k=1}^{K} \theta_k^{Z_{0k} - 1}$$

$$\times \prod_{d=1}^{r} \left\{ \left(\prod_{k=1}^{K} \theta_k^{Z_{dk}} \right) \left(\sum_{k=1}^{K} \theta_k \eta_{dk} \right)^{-\sum_{k=1}^{K} Z_{dk}} \right\} \pi_0(\eta). \tag{17}$$

Next, integrating out ψ, the joint posterior of θ and η is

$$\pi(\theta, \eta | z) \propto \left(\prod_{k=1}^{K} \theta_k^{Z_{0k}} \right) \prod_{d=1}^{r} \left\{ \prod_{k=1}^{K} \left(\frac{\theta_k \eta_{dk}}{\sum_{l=1}^{K} \theta_l \eta_{dl}} \right)^{Z_{dk}} \right\} \prod_{k=1}^{K} \theta_k^{-1} \pi_0(\eta)$$

$$= L_R \left(\prod_{k=1}^{K} \theta_k^{-1} \right) \pi_0(\eta), \tag{18}$$

by (7).

Because both posteriors given in (15) and (18) are derived from the augmented model given in (13), one obtains the desired conclusion that the posterior of η is the same whether it is obtained from the prospective likelihood L_P given in (6) or the retrospective likelihood L_R given in (7).

To verify posterior propriety, the right-hand side of the penultimate line in (18) is bounded above by the function $\prod_{k=1}^{K} \left(\theta_k^{Z_{0k}-1} \right) \pi_0(\eta)$. Because $\pi_0(\eta)$ is proper and $Z_{0k} \geq 1(k = 1, \ldots, K)$, $\int \pi(\theta, \eta | z) \mathrm{d}\theta \mathrm{d}\eta$ is finite.

Appendix B. Proof for posterior equivalence of $(\delta_1, \ldots, \delta_r)$

Note that $\alpha_d = \sum_{l=1}^{K} P_{dl}/P_{0l} = \sum_{l=1}^{K} \exp\{\rho_d + Q(x_l, \delta_d)\}$ so that

$$\log \alpha_d = \rho_d + \log \left[\sum_{l=1}^{K} \exp\{Q(x_l, \delta_d)\} \right], \tag{19}$$

and

$$\eta_{dk} = \alpha_d^{-1}(P_{dk}/P_{0k}) = \frac{\exp\{Q(x_k, \delta_d)\}}{\sum_{l=1}^{K} \exp\{Q(x_l, \delta_d)\}}, \tag{20}$$

which does not depend on α. Independent uniform $(-\infty, \infty)$ priors on the ρ_d $(d = 1, \ldots, r)$ and proper priors on δ_d $(d = 0, \ldots, r)$ result in independent uniform $(-\infty, \infty)$ priors for the $\log \alpha_d$ $(d = 1, \ldots, r)$ or the prior $\prod_{d=1}^{r} \alpha_d^{-1}$ for α and a proper prior for η. Now, the result follows immediately from Theorem 1.

Appendix C. Full conditionals

(1)

$$\pi(\beta_j|\cdot) \propto \phi\left(\frac{\beta_j - \mu_j}{\sigma_j}\right) \prod_{s=1}^{2} \prod_{d=0}^{2} \prod_{k=1}^{4} \left[\frac{\exp(\phi_d x_{kj} \beta_j)}{\sum_{m=0}^{2} \exp(\rho_{sm} + \phi_m x_k^T \beta)}\right]^{Z_{sdk}}$$

where $\mu_j = b_j - \sum_{i \neq j} (\tau_{ji}/\tau_{jj})(\beta_i - b_i)$, $\sigma_j = \sqrt{\tau_{jj}}^{-1}$, $j = 1, 2$. Here, b_i : prior mean of β_i, $i = 1, 2$; τ_{ij} : ij th component of inverse of the prior covariance matrix of β, $i = 1, 2$, $j = 1, 2$; x_{kj} is the j th component of vector x_k, $k = 1, \ldots, 4$, $j = 1, 2$; ϕ is the standard normal PDF.

(2)

$$\pi(\rho_{sd}|\cdot) \propto \frac{\exp\left(\rho_{sd} \sum_{k=1}^{4} Z_{sdk}\right)}{\prod_{d=0}^{2} \prod_{k=1}^{4} \left\{\sum_{m=0}^{2} \exp\left(\rho_{sm} + \phi_m x_k^T \beta\right)\right\}^{Z_{sdk}}}, \qquad s = 1, 2, \quad d = 1, 2.$$

(3)

$$\pi(\phi_1|\cdot) \propto \frac{\exp\left(\phi_1 \sum_{s=1}^{2} \sum_{k=1}^{4} x_k^T \beta Z_{s1k}\right)}{\prod_{s=1}^{2} \prod_{d=0}^{2} \prod_{k=1}^{4} \left\{\sum_{m=0}^{2} \exp\left(\rho_{sm} + \phi_m x_k^T \beta\right)\right\}^{Z_{sdk}}} I(0 \leqslant \phi_1 \leqslant 1).$$

Acknowledgements

References

1. Cornfield J. A method of estimating comparative rates from clinical data: applications to the cancer of the lung, breast and cervix. *Journal of national Cancer Institute* 1951; **11**:1269–1275.
2. Anderson JA. Separate sample logistic discrimination. *Biometrika* 1972; **59**:19–35.
3. Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika* 1979; **66**:403–411.
4. Carroll RJ, Gail MH, Lubin JH. Case-control studies with errors in covariates. *Journal of the American Statistical Association* 1993; **88**:185–199.
5. Roeder K, Carroll RJ, Lindsay BG. A semiparametric mixture approach to case-control studies with errors in covariables. *Journal of the American Statistical Association* 1996; **91**:722–732.
6. Scott AJ, Wild CJ. Fitting logistic models under case-control or choice based sampling. *Journal of the Royal Statistical Society, Series B* 1986; **48**:170–182.
7. Scott AJ, Wild CJ. Fitting logistic models in stratified case-control studies. *Biometrics* 1991; **47**:497–510.
8. Scott AJ, Wild CJ. Fitting regression models to case-control data by maximum likelihood. *Biometrika* 1997; **84**:57–71.
9. Wild CJ. Fitting prospective regression models to case-control data. *Biometrika* 1991; **78**:705–717.
10. Breslow NE, Cain KC. Logistic regression for two-stage case-control data. *Biometrika* 1988; **75**:11–20.
11. Breslow NE, Chatterjee N. Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. *Applied Statistics* 1999; **48**:457–468.
12. Chatterjee N. A two-stage regression model for epidemiological studies with multivariate disease classification data. *Journal of the American Statistical Association* 2004; **99**:127–138.
13. Seaman SR, Richardson S. Equivalence of prospective and retrospective models in the Bayesian analysis of case-control studies. *Biometrika* 2004; **91**:15–25.
14. Ghosh M, Zhang L, Mukherjee B. Equivalence of posteriors in the Bayesian analysis of the multinomial-Poisson transformation. *Metron-International Journal of Statistics* 2006; **LXIV**:19–28.
15. Staicu AM. On the equivalence of prospective and retrospective likelihood methods in case-control studies. *Biometrika* 2010; **97**:990–996.

16. Agresti A. *Categorical Data Analysis*. John Willey & Sons: Hoboken, New Jersey, 2002.
17. Weinberg CR, Wacholder S. Prospective analysis of case-control data under general multiplicative-intercept risk models. *Biometrika* 1993; **80**:461–465.
18. Anderson JA. Regression and ordered categorical variable. *Journal of the Royal Statistical Society, Series B* 1984; **46**:1–40.
19. Greenland S. Alternative models for ordinal logistic regression. *Statistics in Medicine* 1994; **13**:1665–1677.
20. Kuss O. On the estimation of the stereotype regression model. *Computational Statistics and Data Analysis* 2006; **50**:1877–1890.
21. Ahn J, Mukherjee B, Banerjee M, Cooney KA. Bayesian inference for the stereotype regression model: application to a case-control study of prostate cancer. *Statistics in Medicine* 2009; **28**:3139–3157.
22. Ahn J, Mukherjee B, Gruber SB, Sinha S. Missing exposure data in stereotype regression model: application to matched case-control study with disease subclassification. *Biometrics* 2010. DOI: 10.1111/j.1541-0420.2010.01453.x. E-pub ahead of print.
23. Satten GA, Kupper LL. Inferences about exposure-disease associations using probability-of-exposure information. *Journal of the American Statistical Association* 1993; **88**:200–208.
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. Rice KM. Equivalence between conditional and random-effects likelihoods for pair-matched case-control studies. *Journal of the American Statistical Association* 2008; **103**:385–396.
26. Diggle PJ, Morris SE, Wakefield JC. Point-source modeling using matched case control data. *Biostatistics* 2000; **1**:89–105.
27. Ghosh M, Chen M. Bayesian inference for matched case-control studies. *Sankhya, B* 2002; **64**:107–127.
28. Poynter JN, Gruber SB, Higgins PD, *et al.* Statins and the risk of colorectal cancer. *New England Journal of Medicine* 2005; **352**:2184–2192.
29. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society, Series C* 1992; **41**:337–348.
30. Geweke J. Evaluting the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*, Bernardo JM, Smith AFM, Dawid AP, Berger JO (eds). Oxford University Press: Oxford, 1992; 169–193.