Bayesian inference for the stereotype regression model: application to a case–control study of prostate cancer

Jaeil Ahn[1], Bhramar Mukherjee[1, *, †], Mousumi Banerjee[1] and Kathleen A. Cooney[2, 3]

[1]*Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[2]*Department of Internal Medicine and Urology, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[3]*University of Michigan Comprehensive Cancer Center, Ann Arbor, MI 48109, U.S.A.*

SUMMARY

The stereotype regression model for categorical outcomes, proposed by Anderson (*J. Roy. Statist. Soc. B.* 1984; **46**:1–30) is nested between the baseline-category logits and adjacent category logits model with proportional odds structure. The stereotype model is more parsimonious than the ordinary baseline-category (or multinomial logistic) model due to a product representation of the log-odds-ratios in terms of a common parameter corresponding to each predictor and category-specific scores. The model could be used for both ordered and unordered outcomes. For ordered outcomes, the stereotype model allows more flexibility than the popular proportional odds model in capturing highly subjective ordinal scaling which does not result from categorization of a single latent variable, but are inherently multi-dimensional in nature. As pointed out by Greenland (*Statist. Med.* 1994; **13**:1665–1677), an additional advantage of the stereotype model is that it provides unbiased and valid inference under outcome-stratified sampling as in case–control studies. In addition, for matched case–control studies, the stereotype model is amenable to classical conditional likelihood principle, whereas there is no reduction due to sufficiency under the proportional odds model. In spite of these attractive features, the model has been applied less, as there are issues with maximum likelihood estimation and likelihood-based testing approaches due to non-linearity and lack of identifiability of the parameters. We present comprehensive Bayesian inference and model comparison procedure for this class of models as an alternative to the classical frequentist approach. We illustrate our methodology by analyzing data from The Flint Men's Health Study, a case–control study of prostate cancer in African-American men aged 40–79 years. We use clinical staging of prostate cancer in terms of Tumors, Nodes and Metastasis as the categorical response of interest. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: Bayesian inference; case–control study; disease sub-classification

---

*Correspondence to: Bhramar Mukherjee, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.
†E-mail: bhramar@umich.edu

1. INTRODUCTION

In his seminal discussion paper on regression and ordered categorical variables [1], Anderson proposed a very general model for truly discrete outcomes called the *stereotype* regression model. The model can accommodate both ordered and unordered categorical outcomes and allows for inference regarding the order restrictions. Greenland [2] pointed out that unlike models for ordinal data based on the cumulative logits, the stereotype model has the advantage of yielding valid inference under outcome dependent sampling, for example, in case–control studies. In addition to Greenland's observation, in the current paper we note that for matched case–control data with fine degree of stratification and disease sub-classification, the stereotype model allows for the stratum-specific nuisance parameters to be eliminated by using the conditional likelihood principle [3]. The proportional odds model does not yield any reduction due to sufficiency in such instances and the nuisance parameters remain even in the conditional likelihood [4]. However, the stereotype model has been less used, primarily due to some of the computational complexities arising due to non-linearity in the parameters as well as non-standard testing theory for testing the global null hypothesis of no covariate effect [1]. Kuss [5] presents a comprehensive overview of maximum likelihood estimation strategies for the stereotype regression model.

There is a growing literature on parametric and non-parametric Bayesian inference for ordinal data [6–9], but the stereotype model has not received attention in the Bayesian domain so far. A Bayesian approach for inference and model comparison appear to be a natural route for this class of models, and is a new contribution to the literature. Our goal in the current paper is (i) to review the classical background and motivation for proposing this class of models for ordinal data and compare/contrast with other commonly used ordinal models; (ii) point out limitations and issues with maximum likelihood-based inference and illustrate the advantages that an alternative Bayesian route can offer; (iii) describe Bayesian inference for this class of models, with special emphasis on case–control data with finer disease subclassification within the cases; this includes specification of priors, derivation of full conditional distribution of the model parameters and designing the computational algorithm for estimating posterior quantities of interest; (iv) illustrate the methods using a case–control study of prostate cancer among African-American men, conducted in Flint, Michigan. We consider inference under both unmatched and matched case–control designs. Sections 2–5 are organized sequentially according to the above four objectives. We also present a small-scale simulation study to compare the proposed Bayesian method with maximum likelihood based estimation in Section 6. Section 7 presents a brief concluding discussion.

2. CLASSICAL BACKGROUND AND MOTIVATION FOR THE STEREOTYPE REGRESSION MODEL

Two of the popular models for categorical outcomes are the polytomous logit model (or the baseline-category logit model) and the proportional odds model [10, 11]. Let $Y = 0, 1, \ldots, K$ be an outcome with $K+1$ categories and let the reference category be denoted by 0. Let $\mathbf{X}$ be a $p \times 1$ vector of p covariates. Then the proportional odds model is given by, for $k = 0, \ldots, K-1$:

$$\text{logit}[P(Y \leqslant k | \mathbf{X})] = \beta_{0k} - \boldsymbol{\beta}^\top \mathbf{X} \tag{1}$$

$\{\beta_{0k}\}$ are increasing in k and $P(Y \leqslant K | \mathbf{X}) \equiv 1$. The number of parameters to be estimated in this model is $K + p$. The negative sign is used to make positive (negative) values of the elements of $\boldsymbol{\beta}$ correspond to positive (negative) association with the outcome.

On the other hand, the baseline-category logit or polytomous logistic regression model is given by, for $k = 0, 1, \ldots, K$,

$$\Pr(Y = k | \mathbf{X}) = \frac{\exp(\beta_{0k} + \boldsymbol{\beta}_k^\top \mathbf{X})}{\sum_{k=0}^{K} \exp(\beta_{0k} + \boldsymbol{\beta}_k^\top \mathbf{X})} \tag{2}$$

with constraints $\beta_{00} \equiv 0$ and $\boldsymbol{\beta}_0 \equiv \mathbf{0}$. The number of parameters to be estimated in this model is $K + (p \times K)$.

The stereotype model, proposed by Anderson [1] imposes a special structure on the parameters of the polytomous logistic model, namely,

$$\boldsymbol{\beta}_k = \phi_k \boldsymbol{\beta} \tag{3}$$

Thus, the stereotype model in its complete form is represented by

$$\Pr(Y = k | \mathbf{X}) = \frac{\exp(\beta_{0k} + \phi_k \boldsymbol{\beta}^\top \mathbf{X})}{\sum_{k=0}^{K} \exp(\beta_{0k} + \phi_k \boldsymbol{\beta}^\top \mathbf{X})} \tag{4}$$

for $k = 0, 1, \ldots, K$, with $\beta_{00} = \phi_0 \equiv 0$. An additional constraint has to be imposed on the other $\{\phi_k\}$ for identifiability, typically, $\phi_K \equiv 1$. The $\{\phi_k\}$ are regarded as scores for different response categories. Because of this structure, the number of parameters to be estimated in this model are $K + (K - 1) + p = 2K - 1 + p$, in between the dimensionality of models (2) and (1). Note that, akin to the proportional odds model, the stereotype model has the property of representing the effect of the predictors by a single parameter (for given scores) and is more parsimonious than the polytomous logistic model in (2).

In (4), variable X_r has coefficients $\phi_k \beta_r$ (r indexing the variable and k indexing the category) which represents the log odds-ratio corresponding to category k vs category 0 of Y. The constraint $\phi_K \equiv 1$ implies that β_r, corresponding to X_r, represents the effect of unit change in X_r on the log OR of response in the highest category K vs category 0 of Y. In order to understand the interpretation of the parameters in the model and visualize the structure, let us consider the special case with $K = 2$, implying that the possible values of Y are 0, 1, and 2, and a set of two covariates X_1 and X_2. Then the stereotype model and the polytomous logit model (in squared brackets) can be compared as

$$\log \left[\frac{\pi_1(\mathbf{X})}{\pi_0(\mathbf{X})} \right] = \beta_{01} + \phi_1(\beta_1 X_1 + \beta_2 X_2) = [\beta_{01} + \beta_{11} X_1 + \beta_{12} X_2]$$

$$\log \left[\frac{\pi_2(\mathbf{X})}{\pi_0(\mathbf{X})} \right] = \beta_{02} + \phi_2(\beta_1 X_1 + \beta_2 X_2) = [\beta_{02} + \beta_{21} X_1 + \beta_{22} X_2]$$

Just like the proportional odds assumption, the stereotype structure may not be realistic for a given data set, and needs to be tested in terms of model diagnostics and fit statistics. However, it does allow some flexibility compared with the proportional odds model by introducing the category-specific score parameters that are assumed to be the same for each covariate. Allowing covariate-specific scores will lead to a model exactly equivalent to the multinomial logistic model (2).

The stereotype model can also be represented in terms of adjacent category logits

$$\log\left[\frac{\pi_k(\mathbf{X})}{\pi_{k+1}(\mathbf{X})}\right] = \beta_{0k} - v_k\boldsymbol{\beta}^\top\mathbf{X}$$

where $v_k = \phi_{k+1} - \phi_k$ in terms of the score parameters $\{\phi_k\}$'s. With fixed choice of the scores, say equispaced, that is, $\phi_k = k/K$, the model reduces to simpler representations corresponding to the more standard adjacent category logit model as given in [10, p. 287].

Anderson's original motivation was to propose a model for inherently discrete outcomes, different from models derived from grouped-continuous data, like the proportional odds model. The stereotype model captures a situation where the assignment into different categories is done by processing some indeterminate amount of information in the assessor's mind who has 'stereotypes' in which a new case is categorized. Anderson [1] and Greenland [2] both argue that collapsing invariance and invariance under reversal of outcome codes should not be treated as axioms of acceptability for ordered regression models. In some contexts the axiom is less attractive, particularly with *assessed* categorical outcomes.

Anderson [1] suggested using the stereotype model for ordered data with monotone scores:

$$0 \equiv \phi_0 < \phi_1 < \cdots < \phi_K \equiv 1$$

This implies that for a unit increase in covariate x_r, the log(OR) $\phi_k\beta_r$ of category k vs baseline category 0 becomes larger when category k is further away from 0. For this ordered stereotype model the conditional distribution of $Y|\mathbf{X}$ is stochastically ordered in $\boldsymbol{\beta}^\top\mathbf{X}$ [1, Section 3.5]. Note that these orderings are quite different than the ones given in models for grouped continuous data, for example, in the proportional odds model. There, the ordered categories are 'given' and not necessarily ordered with respect to $\mathbf{X}$ and the ordering exists even in the absence of any covariates $\mathbf{X}$. In the stereotype model, the ordering is intrinsically defined through the regression relationships between Y and $\mathbf{X}$. For example, if we had a single binary covariate X with values 0 and 1, and $p_{kl} = \text{pr}(Y = k, X = l)$, $k = 0, \ldots, K$, $l = 0, 1$, denote the cell probabilities in a $2 \times (K+1)$ contingency table, then the observations follow an *ordered stereotype* model if and only if the probability ratios p_{k0}/p_{k1} are monotonically increasing or decreasing. In the spirit of Cornfield's [12] construction of the logistic regression model in terms of the retrospective distribution of $\mathbf{X}|Y = k$, suppose that $\mathbf{X}|Y = k \sim \text{N}(\boldsymbol{\mu}_k, \Sigma)$, $k = 0, \ldots, K$; then the stereotype model implies that $(\boldsymbol{\mu}_k - \boldsymbol{\mu}_0)'\Sigma^{-1} = \phi_k\boldsymbol{\beta}$, or that the means follow a linear trend.

Before concluding this section we will like to point out two major advantages of the stereotype regression model when applied to the case–control data with finer disease sub-classification within the cases, as in our real data example. First, unlike the cumulative logit models, the stereotype model is preserved under outcome dependent sampling. If $f_0, \ldots, f_K$ are the sampling fractions corresponding to each category then, under outcome-stratified sampling,

$$\frac{P(Y = k|\mathbf{X}, \text{sampled})}{P(Y = 0|\mathbf{X}, \text{sampled})} = \exp(\beta_{0k}^* + \phi_k\boldsymbol{\beta}^\top\mathbf{X})$$

where $\beta_{0k}^* = \beta_{0k} + \log(f_k/f_0)$. Thus, under a correctly specified stereotype model the estimates from retrospectively sampled data will be consistent for true covariate effects [2].

Second, in a matched case–control study with binary outcomes, the matched set-specific nuisance parameters are eliminated by using conditional logistic regression [3]. Mukherjee *et al.* [4, 13]

illustrate that the conditioning principle does not apply to the cumulative logit model and the stratum specific parameters remain in the conditional likelihood. The number of such nuisance parameters grows with sample size and thus maximum likelihood estimation runs into inconsistency problems. Owing to its multiplicative intercept structure, the stereotype model is amenable to the conditioning principle. With this backdrop in mind, we proceed to discuss inferential issues associated with this model.

3. INFERENCE IN STEREOTYPE REGRESSION MODEL: ISSUES WITH MAXIMUM LIKELIHOOD AND ADVANTAGES OF A BAYESIAN PARADIGM

In spite of its parsimony when compared to (2), with unknown $\{\phi_k\}$, the non-linearity of the parameters in (4) makes estimation by a standard generalized linear model software infeasible. Greenland [2] suggests a two-step estimation procedure that starts by fixing $\{\phi_k\}$, and then estimating $\boldsymbol{\beta}$ by fitting a baseline-category logit model with multiple predictors $\phi_k x_r$'s. At the second step, treat $\hat{\boldsymbol{\beta}}^\top \mathbf{X}$ as a single predictor and estimate $\{\phi_k\}$ by the same model fitting procedure. This iterative procedure does not guarantee convergence to the true MLE. The correct standard errors are recommended to be obtained by a subsequent bootstrap procedure [2, 14]. Kuss [5] discusses various computational methods to obtain the MLEs in the stereotype regression model. Among the algorithms proposed he discusses generalized least squares (GLS) and a quasi-Newton method for direct maximization of the likelihood. PROC NLMIXED, PROC MODEL in SAS and R function *gnm* can be modified to do the direct maximization. The model can be embedded in the class of reduced rank vector generalized linear models (RR-VGLM) and can be fitted in R-package VGLM [15]. The ordered version of the stereotype model is often harder to fit due to multiple constraints and hitting boundaries of the parameter space and often not implementable in standard software.

A Bayesian procedure can potentially handle many of these difficulties associated with frequentist estimation. The ordering constraints are ensured by appropriate prior choices on the simplex $0 \equiv \phi_0 < \phi_1 < \cdots < \phi_K = 1$, so that the constraints are automatically satisfied while generating posterior samples. Bayesian computation is based on exact simulation of the posterior distribution of the parameters, thus calculating the posterior variance or highest posterior density (HPD) credible interval does not pose any additional challenges or require validity of large sample approximations. Moreover, if one is actually interested in estimating the log-odds-ratio parameters $\phi_k \beta_r$, corresponding to category k and covariate X_r, the Bayesian procedure can directly generate the posterior distribution of this quantity, whereas frequentist inference will require using the multivariate Delta theorem to derive a variance approximation for this product parameter, and rely on large samples in each response category for the validity of this approximation.

The multiplicative nature of the stereotype model poses some issues for testing the null hypothesis of independence $H_0: \boldsymbol{\beta} = 0$ in the likelihood-based framework. Under this global null hypothesis, the score parameters $\{\phi_k\}$ are not identifiable. McCullagh [16] pointed out in the discussion of [1] that the approximate null distribution of the likelihood-ratio statistic is that of the largest eigenvalue of a Wishart matrix [17]. The testing problem under such non-regular conditions remains to be explored in the frequentist domain for this class of models. Theories developed for modified partial likelihood ratio test [18, 19] could be useful in deriving a suitable test under this class of models.

We relegate all the discussion related to hypotheses testing to the supplementary material[‡] and focus on estimation and model comparison in the main text of the paper.

Model comparison between the different class of nested models as well as between ordered and unordered model can be carried out by using the deviance information criterion (DIC) [20, 21]. In a Bayesian numerical integration scheme with direct generation of posterior observations, the expected deviance function $E_{\boldsymbol{\theta}|y}(D(\boldsymbol{\theta}))$ can be estimated by the average deviances over the posterior realizations $\{\boldsymbol{\theta}_l : l = 1, \ldots, L\}$, expressed as

$$\overline{D} = \frac{1}{L} \sum_{l=1}^{L} D(\boldsymbol{\theta}_l) = \frac{1}{L} \sum_{n=1}^{L} \{-2 \log p(y|\boldsymbol{\theta}_l)\}$$

The DIC is defined as the expected deviance, penalized by the effective sample size $\mathrm{DIC} = \overline{D} + (\overline{D} - D(\boldsymbol{\theta}^*)) = 2\overline{D} - D(\boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}^*$ is the posterior mean. The advantage of DIC over the other model selection criteria is that it can be easily computed in a Markov chain Monte Carlo (MCMC) setting, in terms of the readily available posterior quantities.

We now proceed to describe our Bayesian approach with specification of prior, likelihood, posterior and outline the numerical scheme to generate observations from the posterior distribution of model parameters.

4. BAYESIAN INFERENCE FOR THE STEREOTYPE REGRESSION MODEL

We specifically formulate our notations to represent data from a case–control study though the Bayesian proposal is very general and can readily be applied to a prospective stereotype regression model. We first describe the methods for a matched case–control design and then proceed to discuss the modifications under an unmatched study design. Let Y_{ij} denote the disease state corresponding to the jth subject in the ith stratum (or matched set), with S_i denoting all variables that contributed explicitly or implicitly to the formation of the ith stratum. The disease states are classified into one of the K distinct categories $1, 2, \ldots, K$, while the reference control group is denoted by $Y_{ij} = 0$. In each of the N strata we assume that there is one case matched with M controls. The results could be directly generalized to the setting of more general $L_i : M_i$ matching ratio. We consider a vector of covariates $\mathbf{X}_{ij}$ corresponding to subject j in stratum i. The disease risk model is described as

$$\Pr(Y_{ij} = k | \mathbf{X}_{ij}, S_i) = \frac{\exp(\beta_{0k}(S_i) + \phi_k(\boldsymbol{\beta}^\top X_{ij}))}{\sum_{k=0}^{K} \exp(\beta_{0k}(S_i) + \phi_k(\boldsymbol{\beta}^\top X_{ij}))} \quad \text{for } k = 0, \ldots, K \tag{5}$$

The $\beta_{0k}(S_i)$ are category-specific intercepts that could vary with strata. For identifiability $\beta_{00}(S_i) \equiv 0$. Assuming $\beta_{0k}(S_i) \equiv \beta_{0k}$ renders an unmatched analysis where the category-specific intercepts are assumed to be constant across matched sets. Under the ordered model we assume the constraint $0 \equiv \phi_0 < \phi_1 < \cdots < \phi_K \equiv 1$. Without loss of generality, let us assume that the first subject in each stratum is the case and remaining are controls. To eliminate the stratum-specific

---

[‡]Supplementary material associated with this article are available online.

nuisance parameters $\beta_{0k}(S_i)$ we use the conditional likelihood, by conditioning on the event $\sum_{j=1}^{M+1} Y_{ij} = k_i$, in the ith stratum, where k_i is the observed disease state corresponding to the case subject in the ith stratum, that is, $Y_{i1} = k_i$. The corresponding conditional likelihood is given by

$$L_c = \prod_{i=1}^{N} P\left(Y_{i1} = k_i, Y_{i2} = \cdots = Y_{iM+1} = 0 | \{\mathbf{X}_{ij}\}_{j=1}^{M+1}, S_i, \sum_{j=1}^{M+1} Y_{ij} = k_i\right)$$

$$= \prod_{i=1}^{N} \left[\frac{\exp(\phi_{k_i}(\boldsymbol{\beta}^\top \mathbf{X}_{i1}))}{\sum_{j=1}^{M+1} \exp(\phi_{k_i}(\boldsymbol{\beta}^\top \mathbf{X}_{ij}))}\right] \tag{6}$$

One could proceed with Bayesian inference by either assuming a hierarchical prior on the parameters of the prospective model (5) (including $\beta_{0k}(S_i)$) or via the conditional likelihood in (6) treating it as a genuine likelihood and impose prior structure on the parameters ϕ_k and $\boldsymbol{\beta}$. The justification for using L_c as a basis of Bayesian inference can be found in [22].

Priors: There are two sets of parameters under consideration here, the covariate effects $\boldsymbol{\beta}$ and the category-specific scores $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{K-1})$ in the conditional likelihood (6). On the parameters of $\boldsymbol{\beta}$, we impose either independent normal or joint multivariate normal priors. In an *unordered* stereotype model we can similarly put independent or joint multivariate normal prior on $\phi_1, \ldots, \phi_{K-1}$, without an order constraint on the prior support.

For the *ordered* stereotype model, to handle the identifiability condition and the order restriction of the parameters $0 \equiv \phi_0 < \phi_1 < \cdots < \phi_{K-1} < \phi_K = 1$ in a natural way, we reparameterize the parameters in terms of the differences,

$$\gamma_1 = \phi_1, \quad \gamma_s = \phi_s - \phi_{s-1} \quad \text{for } s = 2, \ldots, K-1$$

Thus $\phi_k = \sum_{s=1}^{k} \gamma_s$ for $s = 2, \ldots, K$, and $\phi_0 = \gamma_0 \equiv 0$. The condition $\phi_K \equiv 1$ implies that $\gamma_K = 1 - \phi_{K-1} = 1 - \sum_{s=0}^{K-1} \gamma_s$. The stereotype model requires $\{\phi_k\}$ to be increasing in k and be bounded by 1, this implies that the sequence of parameters $\{\gamma_s\}$ are positive and lie in the $K-1$ dimensional simplex $0 < \sum_{s=1}^{K-1} \gamma_s < 1$. There are several strategies to ensure this; here, we focus on the prior structure involving a Dirichlet distribution as follows:

- If $0 \equiv \phi_0 < \phi_1 < \cdots < \phi_K = 1$ are $K-1$ order statistics from the uniform distribution $U(0, 1)$, then the successive differences of the order statistics, namely, $\{\gamma_s\}$ as defined above, $s = 1, \ldots, K$, follow a Dirichlet$(1, \ldots, 1)$ distribution. This particular prior actually corresponds to centering your belief around equal spacing of the $\{\phi_k\}$. Generalizing this result, one can directly impose a Dirichlet$(\alpha_1, \ldots, \alpha_K)$ distribution jointly on the parameters $\{\gamma_s\}$, which is a natural prior with support as the above simplex.

Full conditional distributions: We first consider the *unordered* stereotype model with the following mutually independent normal priors on $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$:

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \sigma_\beta^2 I_p)$$

$$\boldsymbol{\phi} \sim N(\boldsymbol{\mu}_\phi, \sigma_\phi^2 I_{K-1})$$

Combining the conditional likelihood (6) with the assumed priors, we can derive the following joint posterior distribution:

$$p(\boldsymbol{\beta}, \boldsymbol{\phi}|\mathbf{Y}, \mathbf{X}, \mathbf{S}) \propto \prod_{i=1}^{N} \left[\frac{\exp(\phi_{k_i}(\boldsymbol{\beta}^\top \mathbf{X}_{i1}))}{\sum_{j=1}^{M+1} \exp(\phi_{k_i}(\boldsymbol{\beta}^\top \mathbf{X}_{ij}))} \right] \times \pi(\boldsymbol{\beta})\pi(\boldsymbol{\phi}) \tag{7}$$

The full conditional distribution of the respective parameters are given by

$$\pi(\beta_r|\cdot) \propto \frac{\exp\left(-\frac{1}{2\sigma_{\beta_r}^2}(\beta_r - \mu_{\beta_r} - \sigma_{\beta_r}^2 \sum_{i=1}^{N} \phi_{k_i} X_{i1r})^2\right)}{\prod_{i=1}^{N} \sum_{j=1}^{M+1} \exp(\phi_{k_i}(\boldsymbol{\beta}^\top \mathbf{X}_{ij}))}$$

$$\pi(\phi_k|\cdot) \propto \frac{\exp\left(-\frac{1}{2\sigma_{\phi_k}^2}(\phi_k - \mu_{\phi_k} - \sigma_{\phi_k}^2 \sum_{i=1}^{N} I(Y_{i1}=k)(\boldsymbol{\beta}^\top \mathbf{X}_{ij}))^2\right)}{\prod_{i=1}^{N} \sum_{j=1}^{M+1} \exp \phi_k(\boldsymbol{\beta}^\top \mathbf{X}_{ij})}$$

where X_{ijr} is the value of the rth predictor corresponding to the covariate vector $\mathbf{X}_{ij}$, for the jth subject, in the ith stratum, and β_r is the parameter specific to covariate X_r; $i = 1, \ldots, N$, $j = 1, \ldots, M+1$, $r = 1, \ldots, p$. For the ordered stereotype model, we consider the natural prior choice that $\{\gamma_s\}$, $s = 1, \ldots, K$ follows a Dirichlet$(\alpha_1, \ldots, \alpha_K)$ distribution. The full conditional for $\{\gamma_s\}$ is expressed as

$$\pi(\gamma_s|\cdot) \propto \prod_{i=1}^{N} \left[\frac{\exp(\gamma_s I(Y_{i1} \geqslant s)(\boldsymbol{\beta}^\top \mathbf{X}_{i1}))}{\sum_{j=1}^{M+1} \exp(\sum_{k=1}^{k_i} \gamma_k (\boldsymbol{\beta}^\top \mathbf{X}_{ij}))} \right] \times \gamma_s^{\alpha_s - 1}$$

For the special case of Dirichlet$(1, \ldots, 1)$ prior, corresponding to a prior belief of equal spacing of the scores, the joint full conditional distribution of $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{(K-1)})$ is expressed as follows:

$$\pi(\boldsymbol{\phi}|\cdot) \propto \prod_{i=1}^{N} \left[\frac{\exp(\phi_{k_i}(\boldsymbol{\beta}^\top \mathbf{X}_{i1}))}{\sum_{j=1}^{M+1} \exp(\phi_{k_i}(\boldsymbol{\beta}^\top \mathbf{X}_{ij}))} \right]$$

Remark 1

For *unmatched* case–control data with N subjects consisting of n_1 cases and n_0 controls, we have an intercept parameter that is constant across strata, namely, $\beta_{0k}(S_i) \equiv \beta_{0k}$ and the conditional likelihood need not be evoked, one is able to proceed with the unconditional prospective likelihood (5) with an additional set of priors on $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0K})$. The data vector now has only one index with $(Y_i, \mathbf{X}_i)$ corresponding to the observations on the ith subject, $i = 1, \ldots, N$. For the *unordered* stereotype model, we assume that

$$\boldsymbol{\beta}_0 \sim N(\boldsymbol{\mu}_{\beta_0}, \sigma_{\beta_0}^2 I_K)$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \sigma_\beta^2 I_p)$$

$$\boldsymbol{\phi} \sim N(\boldsymbol{\mu}_\phi, \sigma_\phi^2 I_{K-1})$$

Combining the prospective likelihood with the assumed priors we have the posterior as

$$p(\boldsymbol{\beta}, \boldsymbol{\phi} | \mathbf{Y}, \mathbf{X}) \propto \prod_{i=1}^{N} \left[\frac{\exp(I(Y_i = k)(\beta_{0k} + \phi_k \boldsymbol{\beta}^{\top} \mathbf{X}_i))}{\sum_{k=0}^{K} \exp(\beta_{0k} + \phi_k \boldsymbol{\beta}^{\top} \mathbf{X}_i)} \right] \times \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\phi}) \tag{8}$$

The full conditional distribution of the respective parameters is as follows:

$$\pi(\beta_{0k} | \cdot) \propto \frac{\exp\left(-\frac{1}{2\sigma_{\beta_{0k}}^2} (\beta_{0k} - \mu_{\beta_{0k}} - \sigma_{\beta_{0k}}^2 \sum_{i=1}^{N} I(Y_i = k))^2 \right)}{\prod_{i=1}^{N} \sum_{k=0}^{K} \exp(\beta_{0k} + \phi_k \boldsymbol{\beta}^{\top} \mathbf{X}_i)}$$

$$\pi(\beta_r | \cdot) \propto \frac{\exp\left(-\frac{1}{2\sigma_{\beta_r}^2} (\beta_r - \mu_{\beta_r} - \sigma_{\beta_r}^2 \sum_{i=1}^{N} I(Y_i = k)\phi_k X_{ir})^2 \right)}{\prod_{i=1}^{N} \sum_{k=0}^{K} \exp(\beta_{0k} + \phi_k \boldsymbol{\beta}^{\top} \mathbf{X}_i)}$$

$$\pi(\phi_k | \cdot) \propto \frac{\exp\left(-\frac{1}{2\sigma_{\phi_k}^2} \left(\phi_k - \mu_{\phi_k} - \sigma_{\phi_k}^2 \sum_{i=1}^{N} I(Y_i = k)(\boldsymbol{\beta}^{\top} \mathbf{X}_i) \right)^2 \right)}{\prod_{i=1}^{N} \sum_{k=0}^{K} \exp(\beta_{0k} + \phi_k \boldsymbol{\beta}^{\top} \mathbf{X}_i)}$$

Here X_{ir} is the rth covariate corresponding to subject i, $i = 1, \ldots, N$, $r = 1, \ldots, p$. For the *ordered* stereotype model and unmatched data likelihood, the full conditional distribution of $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$ under the Dirichlet$(\alpha_1, \ldots, \alpha_k)$ prior is expressed as follows:

$$\pi(\gamma_s | \cdot) \propto \prod_{i=1}^{N} \left[\frac{\exp(\gamma_s I(Y_i \geqslant s)(\boldsymbol{\beta}^{\top} \mathbf{X}_i))}{\sum_{k=0}^{K} \exp(\beta_{0k} + \phi_k \boldsymbol{\beta}^{\top} \mathbf{X}_i)} \right] \times \gamma_s^{\alpha_s - 1}$$

Bayesian computation: The Gibbs sampler [23, 24] is used to generate a sequence of random observations from the joint posterior distribution $[\boldsymbol{\beta}, \boldsymbol{\phi} | Y, \mathbf{X}, S]$ and Bayesian estimates are obtained from this generated sequence. In this algorithm, observations are sampled iteratively from the following conditional distributions: $p[\boldsymbol{\beta} | \mathbf{X}, Y, S, \boldsymbol{\phi}]$ and $p[\boldsymbol{\phi} | \mathbf{X}, Y, S, \boldsymbol{\beta}]$. Since none of the full conditionals have a standard distributional form, we used a hybrid of the Gibbs sampler and the Metropolis Hastings algorithm to generate random numbers from the full conditionals of the parameters as specified above. Noting that some of the above full-conditionals are log-concave, we use adaptive rejection sampling (ARS) [25] for sampling $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ for the unordered model and only $\boldsymbol{\beta}$ for the ordered model. We use the Metropolis-Hastings update of the $\boldsymbol{\phi}$ in the ordered model with proposal distribution on ϕ as order statistics from the uniform distribution on $[0, 1]$. We typically ran the sampler 50 000 iterations and considered estimates based on the last 10 000 runs, allowing a burn-in of 40 000 iterations. To monitor the convergence of the chain, we use the 'potential scale reduction factor' diagnostic proposed by Gelman and Rubin [26].

J. AHN *ET AL.*

5. EXAMPLE: THE FLINT MEN'S HEALTH STUDY

Prostate cancer is the most common non-cutaneous cancer among American men and is the second leading cause of cancer deaths in the United States. African American men have a 1.5-fold higher incidence of prostate cancer compared to Caucasian men [27]. The Flint Men's Health Study (FMHS) is a community-based case–control study of prostate cancer in African-American men between the ages of 40–79, from 1999 to 2002 [28]. Case subjects were identified using the Genessee County Community-wide hospital oncology program (CHOP) registry. Along with in-home interview questionnaires on occupational and behavioral exposures, personal and family history of cancer, the study collected hospital record information on stage, Gleason's grade of differentiation, treatment and pre-diagnosis prostate-specific antigen (PSA) value. Several anthropometric measurements and blood sample were also collected. Disease-free controls were identified from a sample of African-American men in Flint, or in selected census tracts in neighboring communities. Controls were asked to undergo a prostate cancer screening protocol, which included providing blood sample for PSA measurement and a comprehensive urological examination. Men with abnormal clinical examination or elevated PSA were excluded from control set and referred for prostate biopsy. There were 28 men who were subsequently diagnosed with prostate cancer from this referral, who were included as case subjects. We had initial data on 144 cases and 434 controls, which contained some missing information.

Our categorical outcome variable is the stage of prostate cancer. Two systems are in common use for the staging of prostate cancer. The Jewett system ([29], stages A through D) was described in 1975 and has since been modified. In 1997, the American Joint Committee on Cancer (AJCC) and the International Union Against Cancer adopted a revised tumor, nodes, metastasis (TNM) system that employs the same broad stage categories as the Jewett system but includes finer subcategories of stage. This revised TNM system is clinically useful and more precisely stratifies newly diagnosed patients. In 2002, the AJCC further revised the TNM classification system [30]. A thorough review of the controversies of staging in prostate cancer is contained in [31]. We have used the AJCC recommended TNM-based classifications to define our response variable. Notice that stage is not a one-dimensional continuum or discretized version of such a latent continuum. According to Anderson's original motivating framework, assignment of stage should be treated as summaries of multidimensional outcomes. The *stereotypes* assigned for different stages are based on many aspects of a clinical and histopathological examination. As two potential predictors, we used $\log(1+\text{PSA})$ and age. We used pre-diagnosis total free PSA for the case subjects and the total free PSA measured at the time of entry in the study for control subjects as a measure of PSA. The age recorded at the time of the PSA measurement is used as a second covariate in the model.

The final data set we use for illustration purposes, after removing subjects with missing information on stage, age or PSA, consists of 433 controls and 132 cases. Among 132 cases, 64 subjects are in stage 1, 61 subjects are in stage 2 and 7 subjects are in stage 3 according to the AJCC/TNM classification as described above. Ages are noted to have a roughly uniform distribution over the entire range [40,79] with mean 58 years and standard deviation 10.4 years. To avoid very small values for the estimated regression coefficient corresponding to age, we transform the ages into a [0, 1] interval by using the linear transformation, that is, (Age-40)/(79–40). The marginal empirical distribution of log-transformed PSA (in combined case and control sample) is distributed roughly as a normal distribution with mean 1.12 and standard deviation 0.84 truncated at 0. We first analyze the unmatched data using the unconditional likelihood (5) with $\beta_{0k}(S_i)=\beta_{0k}$, and explore the association of age and PSA with different stages of cancer. We fit both the ordered and

Copyright © 2009 John Wiley & Sons, Ltd.

Statist. Med. 2009; **28**:3139–3157
DOI: 10.1002/sim

J. AHN *ET AL.*

5. EXAMPLE: THE FLINT MEN'S HEALTH STUDY

Prostate cancer is the most common non-cutaneous cancer among American men and is the second leading cause of cancer deaths in the United States. African American men have a 1.5-fold higher incidence of prostate cancer compared to Caucasian men [27]. The Flint Men's Health Study (FMHS) is a community-based case–control study of prostate cancer in African-American men between the ages of 40–79, from 1999 to 2002 [28]. Case subjects were identified using the Genessee County Community-wide hospital oncology program (CHOP) registry. Along with in-home interview questionnaires on occupational and behavioral exposures, personal and family history of cancer, the study collected hospital record information on stage, Gleason's grade of differentiation, treatment and pre-diagnosis prostate-specific antigen (PSA) value. Several anthropometric measurements and blood sample were also collected. Disease-free controls were identified from a sample of African-American men in Flint, or in selected census tracts in neighboring communities. Controls were asked to undergo a prostate cancer screening protocol, which included providing blood sample for PSA measurement and a comprehensive urological examination. Men with abnormal clinical examination or elevated PSA were excluded from control set and referred for prostate biopsy. There were 28 men who were subsequently diagnosed with prostate cancer from this referral, who were included as case subjects. We had initial data on 144 cases and 434 controls, which contained some missing information.

Our categorical outcome variable is the stage of prostate cancer. Two systems are in common use for the staging of prostate cancer. The Jewett system ([29], stages A through D) was described in 1975 and has since been modified. In 1997, the American Joint Committee on Cancer (AJCC) and the International Union Against Cancer adopted a revised tumor, nodes, metastasis (TNM) system that employs the same broad stage categories as the Jewett system but includes finer subcategories of stage. This revised TNM system is clinically useful and more precisely stratifies newly diagnosed patients. In 2002, the AJCC further revised the TNM classification system [30]. A thorough review of the controversies of staging in prostate cancer is contained in [31]. We have used the AJCC recommended TNM-based classifications to define our response variable. Notice that stage is not a one-dimensional continuum or discretized version of such a latent continuum. According to Anderson's original motivating framework, assignment of stage should be treated as summaries of multidimensional outcomes. The *stereotypes* assigned for different stages are based on many aspects of a clinical and histopathological examination. As two potential predictors, we used $\log(1+\text{PSA})$ and age. We used pre-diagnosis total free PSA for the case subjects and the total free PSA measured at the time of entry in the study for control subjects as a measure of PSA. The age recorded at the time of the PSA measurement is used as a second covariate in the model.

The final data set we use for illustration purposes, after removing subjects with missing information on stage, age or PSA, consists of 433 controls and 132 cases. Among 132 cases, 64 subjects are in stage 1, 61 subjects are in stage 2 and 7 subjects are in stage 3 according to the AJCC/TNM classification as described above. Ages are noted to have a roughly uniform distribution over the entire range [40,79] with mean 58 years and standard deviation 10.4 years. To avoid very small values for the estimated regression coefficient corresponding to age, we transform the ages into a [0, 1] interval by using the linear transformation, that is, (Age-40)/(79–40). The marginal empirical distribution of log-transformed PSA (in combined case and control sample) is distributed roughly as a normal distribution with mean 1.12 and standard deviation 0.84 truncated at 0. We first analyze the unmatched data using the unconditional likelihood (5) with $\beta_{0k}(S_i)=\beta_{0k}$, and explore the association of age and PSA with different stages of cancer. We fit both the ordered and

Copyright © 2009 John Wiley & Sons, Ltd.

Statist. Med. 2009; **28**:3139–3157
DOI: 10.1002/sim

unordered model in the proposed Bayesian framework. We used $N(0, 5^2)$ prior distributions for $\boldsymbol{\beta}, \boldsymbol{\phi}$ in unordered stereotype model and $\boldsymbol{\beta}$ in the ordered model. For $\boldsymbol{\phi}$ in the ordered model, we use the Dirichlet$(1, 1, 1, 1)$ as a prior density for the successive score differences $\{\gamma_s\}$ as described in the previous section.

In order to obtain the MLE under the unordered stereotype regression model, we follow the maximum likelihood algorithm implemented in the SAS/ETS module PROC MODEL, by fitting the stereotype model as a general multivariate non-linear regression model [5]. Throughout the paper we refrain from maximizing the likelihood under the ordered stereotype model due to computational instability. We provide the result of baseline category or multinomial logistic regression model for reference. Though not quite comparable, we present the results of a proportional odds model fitted to the data as well.

Estimates of $\boldsymbol{\beta}, \boldsymbol{\phi}$ and their standard errors with 95 per cent HPD for Bayesian methods or 95 per cent confidence interval for frequentist methods under different models and approaches are presented in Table I. We can note the well expected significant positive association between PSA and stage of cancer. The corresponding log-odds-ratios given in Supplementary Table I also illustrate that the odds-ratios corresponding to each stage as compared with the controls appear to be ordered in terms of PSA. On the other hand, the effect of age, after adjusting for PSA is not so pronounced and the direction is sensitive to the model choice, with the age effect not being significant under any method, except attaining borderline significance in the ML fitted stereotype model. The score parameters $\hat{\boldsymbol{\phi}}$ appear to be monotonically increasing across stage, though their values are quite close. The ordered and unordered Bayesian stereotype models have very similar values of estimated DIC criterion. The unordered stereotype regression model fitted by the approach followed in [5] has the AIC estimate (though directly not comparable to DIC) close to the Bayesian models. The multinomial logistic model has an AIC that is higher than the stereotype model. The variance of the log-odds-ratio corresponding to category 3 is indeed worrisome in the multinomial logistic regression model (estimate 10.81, s.e. $=7.47$). This is due to the fact that there are only seven subjects in category 3 who contribute to the estimation of this parameter, whereas the stereotype model uses common covariate-specific parameter for estimating this log-odds-ratio and has much superior precision. The proportional odds model, on the other hand has a higher AIC value compared to both multinomial logistic and stereotype model, indicating that this model may not be the right choice in this context.

For illustration purposes we created a matched case–control data set from this database with 132 cases matched with three controls in terms of census tract and neighborhoods of residence to obtain a 1:3 matched data set with 396 controls. With these 528 subjects and 132 matched sets, we fitted the stereotype and multinomial logistic model via the conditional likelihood. We omit presenting the results for the proportional odds model for the matched case, as it is not amenable to the conditioning principle and can only implement an unmatched analysis of a matched study.

For the stereotype models for matched data we consider both ordered and unordered models with the same prior distributions for unmatched data analysis under the Bayesian approach. Since the approach in [5] does not discuss conditional likelihood, we obtain the conditional maximum likelihood estimates by direct maximization of (6) via the Nelder–Mead optimization routine. The optimization method though relatively stable for a few parameters, encounters many convergence issues with larger number of covariates or response categories.

The results of matched data analysis are presented in Table II. Though there are certain numerical differences especially, in terms of the effects of age when compared with the unmatched analysis, the basic pattern of inference remains the same. The ordered stereotype model appears to provide

Table I. Results of analysis of unmatched FMHS data with 132 cases and 433 controls.

		β_1	β_2	ϕ_1	ϕ_2	DIC/AIC*
Unordered stereotype						
Bayes	Estimate	−0.12	4.29	0.52	0.57	557.6
	PSD	1.11	0.46	0.06	0.06	
	95% HPD	(−2.31, 1.99)	(3.39, 5.15)	(0.40, 0.64)	(0.46, 0.70)	
MLE	Estimate	−1.92	8.72	0.35	0.39	563.5
	s.e.	0.88	0.82	0.04	0.04	
	95% CI	(−3.64, −0.19)	(7.11, 10.33)	(0.27, 0.43)	(0.31, 0.47)	
		β_1	β_2	ϕ_1	ϕ_2	DIC
Ordered stereotype						
Bayes	Estimate	−0.01	4.44	0.50	0.56	555.9
	PSD	0.99	0.38	0.05	0.05	
	95% HPD	(−2.28, 2.24)	(3.57, 5.24)	(0.40, 0.61)	(0.45, 0.68)	
		β_1	β_2			AIC
Alternative models						
Proportional odds logistic regression	Estimate	0.01	2.13			570.0
	s.e.	0.59	0.17			
	95% CI	(−1.15, 1.17)	(1.79, 2.46)			
		β_{11}	β_{12}	β_{13}		AIC
Multinomial logistic regression[†]	Estimate	1.37	−0.35	−1.08		
	s.e.	0.79	0.83	6.15		
	95% CI	(−0.18, 2.93)	(−1.98, 1.27)	(−13.13, 10.96)		
		β_{21}	β_{22}	β_{23}		567.2
	Estimate	2.19	2.60	10.81		
	s.e.	0.24	0.26	7.47		
	95% CI	(1.71, 2.67)	(2.08, 3.11)	(−3.83, 25.45)		

We used transformed age and $\log(1+\text{PSA})$ as the two covariates X_1 and X_2. Under the Bayesian methods the 'estimate' corresponds to the posterior mean, whereas PSD and HPD correspond to posterior standard deviation and highest posterior density intervals, respectively. For the MLE, CI corresponds to the Wald-type large sample confidence intervals.
*AIC is the akaike information criterion and DIC is the deviance information criterion.
[†]In multinomial logistic regression model, β_{rk} is the log-odds-ratio of category k vs the controls corresponding to covariate X_r, $r=1,2$; $k=1,2,3$.

a slightly better fit to the data in terms of the DIC criterion. The posterior standard deviations are larger for the matched data set, providing wider credible intervals. This may be due to the reason that we created artificial matching in the data set for illustration purposes and the matching strategy may not have been efficient and created correlation among covariates within a matched set (Table I).

Figures 1 and 2 show posterior densities of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ for the ordered and the unordered model, overlaid on each other for unmatched and matched data sets, respectively. The ordered and unordered

Table II. Results of analysis of 1:3 matched FMHS data with transformed age and
$\log(1+\text{PSA})$ as the two covariates X_1 and X_2.

		β_1	β_2	ϕ_1	ϕ_2	DIC/AIC*
Unordered stereotype						
Bayes	Estimate	1.57	4.56	0.54	0.63	145.6
	PSD	1.15	1.67	0.20	0.23	
	95% HPD	$(-0.72, 4.67)$	$(1.95, 8.59)$	$(0.20, 1.03)$	$(0.26, 1.13)$	
MLE	Estimate	1.54	4.37	0.51	0.58	148.6
	s.e.	1.33	1.78	0.20	0.23	
	95% CI	$(-1.07, 4.15)$	$(0.88, 7.86)$	$(0.12, 0.90)$	$(0.13, 1.03)$	
		β_1	β_2	ϕ_1	ϕ_2	DIC
Ordered stereotype						
Bayes	Estimate	1.57	4.33	0.50	0.66	144.3
	PSD	1.18	1.08	0.13	0.16	
	95% HPD	$(-0.73, 3.96)$	$(2.45, 6.50)$	$(0.29, 0.80)$	$(0.40, 0.96)$	
		β_{11}	β_{12}	β_{13}		AIC
Alternative models						
Multinomial	Estimate	1.30	0.29	—		
conditional	s.e.	0.84	0.99	—		
logistic[†]	95% CI	$(-0.33, 2.94)$	$(-1.66, 2.24)$	—		
		β_{21}	β_{22}	β_{23}		—
	Estimate	2.09	2.61	—		
	s.e.	0.41	0.45	—		
	95% CI	$(1.27, 2.90)$	$(1.73, 3.49)$	—		

There are 132 matched sets. Under the Bayesian methods the 'estimate' corresponds to the posterior mean, whereas PSD and HPD correspond to posterior standard deviation and highest posterior density intervals, respectively. For the MLE, CI corresponds to the Wald-type large sample confidence intervals. For Bayesian models we report DIC, whereas for models fitted by maximum likelihood we report AIC.
*AIC is the akaike information criterion and DIC is the deviance information criterion.
[†]The multinomial logistic model for matched data was fitted by simultaneously maximizing three conditional likelihoods of categories 1 vs 0, 2 vs 0 and 3 vs 0. The conditional MLEs of β_{13}, β_{23} did not converge, thus the AIC measure is not available.

models produce fairly similar results in the unmatched case (Figure 1), but certain differences are noted in the matched data context (Figure 2). For the ordered model, mixing of the Markov chain is slow with a single Dirichlet distribution as proposal density for the ϕ. We used the mixture of two Dirichlet distributions as proposal and increased the number of iterations to 100 000 for the ordered models, however, certain roughness is still noted in the density plot due to ties in the posterior samples at boundaries.

An advantage of the Bayesian approach is to be able to carry out exact posterior inference on the log-odds-ratio parameters of category k compared to category 0, namely $\phi_k \beta_r$, $k=1,2,3$ and $r=1,2$, corresponding to covariate X_r. Supplementary Tables I and II present the corresponding results for the unmatched and matched data sets. Supplementary Figures 1 and 2 exhibit the
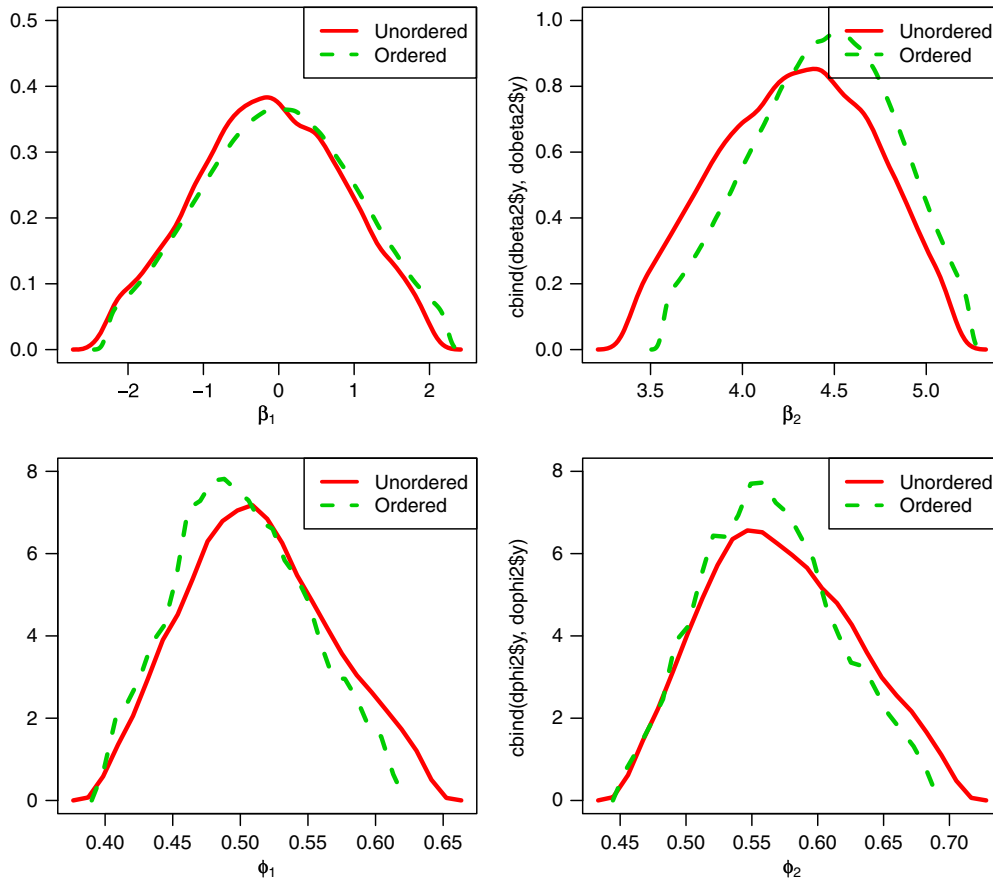
Figure 1. Posterior density estimates for covariate and category-specific parameters of the stereotype model for unmatched FMHS data with numerical summaries as presented in Table I. The results are based on 100 000 observations generated from the posterior distribution of each parameter. The solid line corresponds to the unordered model, whereas the dashed line corresponds to the ordered model.

posterior distribution of the log-odds-ratio parameters $\phi_k \beta_r$, $r = 1, 2$, $k = 1, 2, 3$; which look fairly symmetric for the unmatched data but exhibit skewness in matched data. Based on our computational experience, it appears that both likelihood maximization and Bayesian estimation with the conditional likelihood for matched data are appreciably more challenging than the unconditional likelihood with stereotype link function.

6. SIMULATION STUDY

As in a real data set we can only illustrate the methods and the truth about the parameters is unknown, we conducted a small-scale simulation study to evaluate our proposed methods, mimicking the real data analysis of FMHS in certain aspects. We consider two covariates, namely,
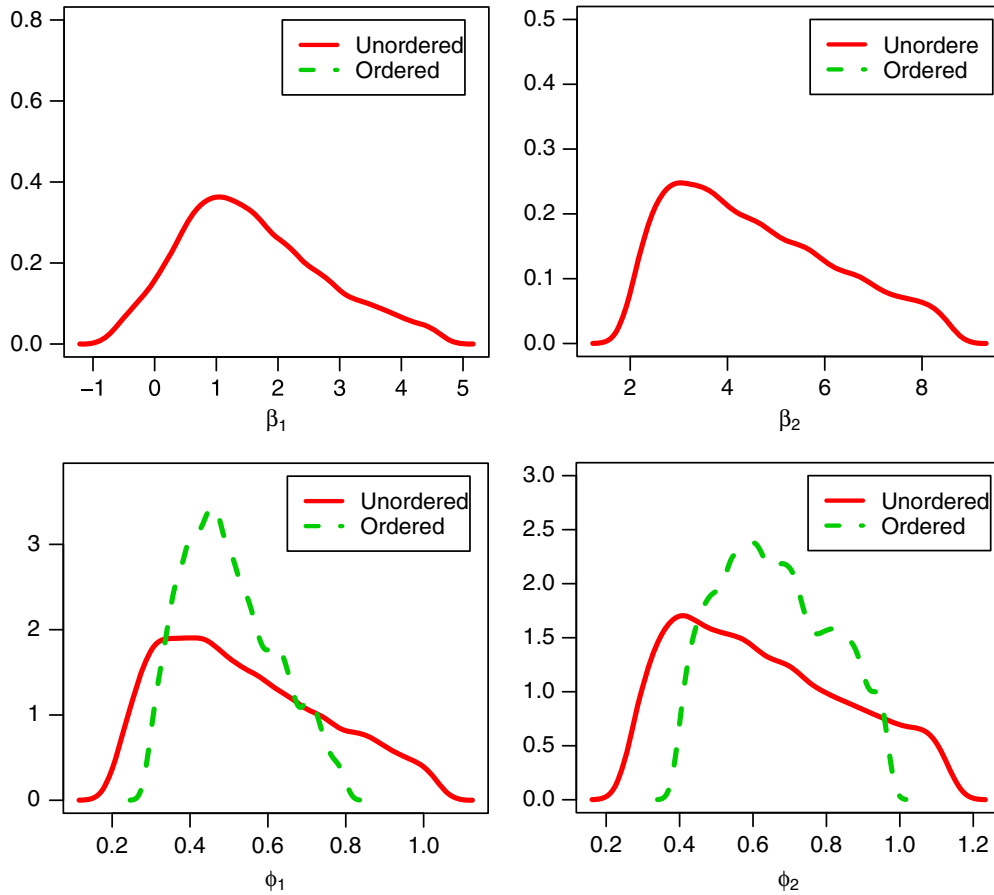
Figure 2. Posterior density estimates for covariate and category-specific parameters of the stereotype model for 1:3 matched FMHS data with numerical summaries as presented in Table II. The results are based on 10 000 observations generated from the posterior distribution of each parameter. The solid line corresponds to the unordered model, whereas the dashed line corresponds to the ordered model.

X_1 and X_2 resembling age and PSA in the FMHS study. We generate X_1 from a uniform distribution on [40,79] and $\log(X_2)$ from a N(1.12, 0.84^2) distribution. We then generate a matching variable S following a U[0, 1] distribution, which will only be used for generating the matched sample. Given the true parameter values and $\mathbf{X}$, we generate the outcome variable $Y|\mathbf{X}, S$ with five categories ranging from 0 to 4 from a multinomial distribution with the following response probability:

$$\Pr(Y_i = k|\mathbf{X}_i, S_i) = \frac{\exp(\beta_{0k} + \beta_S S_i + \phi_k(\boldsymbol{\beta}^\top \mathbf{X}_i))}{\sum_{k=0}^{K} \exp(\beta_{0k} + \beta_s S_i + \phi_k(\boldsymbol{\beta}^\top \mathbf{X}_i))} \quad \text{for } k = 0, \dots, 4$$

where β_{0k} are category-specific intercepts with $\beta_{00} = 0$. We set the effect size corresponding to the matching variable, namely, β_S at 0.5. We consider the covariate-specific parameters $(\beta_1, \beta_2) = (1.0, 2.0)$. We consider two scenarios with respect to the category-specific scores: $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \phi_3, \phi_4) = (0, 0.6, 0.9, 0.6, 1)$ reflecting that the ordering assumption is violated

and $\phi = (0, 0.25, 0.5, 0.75, 1)$ when the ordering assumption is preserved. We call these two situations unordered and ordered setting, respectively. The category-specific intercepts β_{0k} are set as $(0, -5.5, -7, -6, -8)$ for unordered setting and $(0, -4.5, -5.5, -6.5, -8)$ for ordered setting.

With the above distributional structure, we generate 500 000 independent realizations of the data vector $(Y, \mathbf{X}, S)$ resulting in a cohort with 500 000 subjects with roughly 92 per cent controls and 8 per cent cases. The percentage of cases in stages 1, 2, 3, and 4 were 45, 30, 15 and 10 per cent, respectively. We first randomly select 150 cases from the case population. In order to construct a 1:3 matched data set, corresponding to each selected case we selected three controls having the value of the matching variable S within 0.03 of what was noted for the case subject. For generating the unmatched data sets we simply remove S from the above simulation scheme and generate $Y|\mathbf{X}$ only, and randomly select 150 cases and 450 controls from case and control population, respectively. We generated 500 such matched and unmatched data sets by sampling from the cohort. Thus, we have four simulation settings (a) unmatched, unordered (b) unmatched, ordered (c) matched, unordered and (d) matched, ordered as presented in Tables III and IV, respectively.

We fitted both the ordered and unordered stereotype models using the Bayesian approach, with priors identical to our data analysis section. For the unordered model we consider maximum likelihood estimation as well. With unmatched, unordered data in Table III we notice that the Bayesian ordered stereotype model yields larger biases with respect to ϕ though the estimation of β is quite satisfactory in terms of both bias and MSE. The unordered stereotype model as implemented by the methods illustrated in [5] encounters convergence problems in roughly 30 per cent of all

Table III. The results of the simulation study under unmatched case–control sampling design.

		β_1	β_2	ϕ_1	ϕ_2	ϕ_3
Unordered setting						
Unordered stereotype	bias	-0.02	-0.06	0.04	0.03	0.06
Bayes	MSE	0.25	0.08	0.02	0.03	0.02
Unordered stereotype	bias	0.20	0.14	0.04	-0.00	0.03
MLE	MSE	0.38	0.96	0.03	0.05	0.03
Ordered stereotype	bias	0.05	0.08	-0.07	-0.18	0.18
Bayes	MSE	0.33	0.08	0.01	0.04	0.04
Ordered setting						
Unordered stereotype	bias	-0.12	0.14	0.01	0.02	0.03
Bayes	MSE	0.30	0.12	0.01	0.10	0.15
Unordered stereotype	bias	0.12	0.13	-0.01	-0.00	-0.01
MLE	MSE	0.59	0.25	0.01	0.03	0.04
Ordered stereotype	bias	-0.08	-0.08	0.00	0.00	0.00
Bayes	MSE	0.29	0.06	0.01	0.01	0.01

The estimated DICs corresponding to Bayesian estimation of the unordered stereotype model and ordered stereotype model are 922.8 and 928.1 and for unordered setting, 973.5 and 968.1 for the ordered setting, respectively.

The results are based on 500 simulated data sets with 150 cases and 450 controls. For each parameter we report estimated bias and mean-squared error based on the 500 replications. The outcome variable Y has five categories from 0 to 4. The true values for the parameters are: $\beta_1 = 1.0$, $\beta_2 = 2.0$; for unordered setting $\phi_1 = 0.6$, $\phi_2 = 0.9$, $\phi_3 = 0.6$ and for ordered setting $\phi_1 = 0.25$, $\phi_2 = 0.5$ and $\phi_3 = 0.75$.

Table IV. The results of the simulation study under matched case–control sampling design with 1:3 matching ratio.

		β_1	β_2	ϕ_1	ϕ_2	ϕ_3
Unordered setting						
Unordered stereotype	bias	0.00	0.01	0.09	0.14	0.15
Bayes	MSE	0.38	0.30	0.08	0.12	0.13
Unordered stereotype	bias	0.15	0.30	0.01	0.01	0.05
MLE	MSE	0.77	1.02	0.06	0.10	0.11
Ordered stereotype	bias	0.22	0.43	−0.15	−0.26	0.15
Bayes	MSE	0.46	0.34	0.03	0.07	0.03
Ordered setting						
Unordered stereotype	bias	0.01	−0.01	0.04	0.07	0.15
Bayes	MSE	0.39	0.27	0.02	0.05	0.15
Unordered stereotype	bias	0.13	0.23	0.01	0.00	0.04
MLE	MSE	0.56	0.39	0.02	0.04	0.12
Ordered stereotype	bias	0.12	0.17	−0.01	−0.02	−0.01
Bayes	MSE	0.43	0.17	0.01	0.01	0.01

The estimated DICs for unordered stereotype Bayes model and ordered stereotype Bayes model are 281.8 and 285.3 and for unordered setting, and, 320.3 and 315.4 for the ordered setting, respectively.
The results are based on 500 simulated data sets, each with 150 matched sets. For each parameter we report estimated bias and mean squared error based on the 500 replications. The outcome variable Y has five categories from 0 to 4. The true values for the parameters are: $\beta_1 = 1.0$, $\beta_2 = 2.0$; for unordered setting $\phi_1 = 0.6$, $\phi_2 = 0.9$, $\phi_3 = 0.6$ and for ordered setting $\phi_1 = 0.25$, $\phi_2 = 0.5$ and $\phi_3 = 0.75$.

generated data sets and even after removing the non-convergent data sets the MLE have larger bias and MSE than the corresponding Bayesian method for estimating $\boldsymbol{\beta}$. Under (a), the unordered Bayesian method performs best when estimations of $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ are considered simultaneously. In contrast, with (b) unmatched ordered data, the ordered stereotype model as implemented by the Bayesian method performs the best in terms of mean-squared errors for both $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. Note that ML approach tends to yield smaller biases in estimating $\boldsymbol{\phi}$ while producing larger biases in estimating $\boldsymbol{\beta}$, an observation that we could not explain theoretically or heuristically. Interestingly the average DIC for unordered (ordered) model is less for unordered (ordered) setting, thus the two simulation settings are reflected in the preferred model choice under the Bayesian approaches.

We repeat the simulation with 500 *matched* data sets and results are given in Table IV. The basic pattern remains the same as in Table III, with the Bayesian unordered model performing the best under (c) and the Bayesian ordered model performing the best under (d). Once again, the DICs are able to indicate the better model fit among the ordered and unordered model, consistent with the simulation setting of generating ordered or unordered data.

Supplementary Tables III and IV contain the corresponding results for the log-odds-ratio parameters and the same basic pattern is followed. Based on the simulation study and the issues encountered with convergence of MLE, it appears that fitting the ordered and unordered stereotype regression model by the Bayesian approach and assessing comparative model fit via Bayesian model fitting diagnostics is an attractive alternative for this class of models.

7. DISCUSSION

The *stereotype* regression model is an interesting class of models for categorical outcomes which has remained somewhat unexplored in the literature due to problems with classical inference as discussed in Section 3. The Bayesian paradigm circumvents many of the issues with classical inference and provides an alternative approach to estimate the model parameters and carry out model comparisons and model selection for this class of models. In modern medicine, characterization of disease subclasses in terms of histological and morphological terms is often available. The *stereotype* model has a unique distinction among models for ordered data that is preserved under outcome-dependent sampling and can be applied to matched data with fine degree of stratification. The current paper is the first attempt toward proposing Bayesian inference for this class of models with and without the ordering restrictions. We also aim to provide the reader with a classical overview of this class of models. We point out its advantages for unmatched and matched case–control designs and illustrate the methodology in a study of prostate cancer. We present a simulation study when the true outcomes are generated from both ordered and unordered model and illustrate that the Bayesian model comparison approach can discern between the two situations and lead to an estimate with good mean-squared error properties.

There are many issues that need further exploration related to the stereotype model from a frequentist analysis perspective as well, among which is the exploration of an appropriate testing strategy for the hypothesis of independence. Our study indicates that we need better computation strategies in the conditional likelihood setting. How to extend this model to accommodate missing data, correlated or clustered observations can be considered as topics of future research. The codes for maximum likelihood estimation and Bayesian estimation are available at http://www.sph.umich.edu/bhramar/public_html/research.

REFERENCES

1. Anderson JA. Regression and ordered categorical variable. *Journal of Royal Statistical Society*, *Series B* 1984; **46**:1–30.
2. Greenland S. Alternative models for ordinal logistic regression. *Statistics in Medicine* 1994; **13**:1665–1677.
3. Breslow NE, Day NE. *Statistical Methods in Cancer Research. Vol. 1—The Analysis of Case-Control Studies*. IARC Scientific Publications: Lyon, France, 1980.
4. Mukherjee B, Liu I, Sinha S. Analysis of matched case–control data with ordinal disease states: possible choices and comparisons. *Statistics in Medicine* 2007; **26**(17):3240–3257.
5. Kuss O. On the estimation of the stereotype regression model. *Computational Statistics and Data Analysis* 2006; **50**:1877–1890.
6. Johnson VE, Albert JH. *Ordinal Data Modeling*. Springer: Berlin, 1999.
7. Congdon P. *Bayesian Models for Categorical Data*. Wiley: New York, 2005.
8. Kottas A, Müller P, Quintana F. Nonparametric Bayesian modeling for multivariate ordinal data. *Journal of Computational and Graphical Statistics* 2005; **14**:610–625.

9. Sinha S, Mukherjee B, Ghosh M. Bayesian semiparametric modeling for matched case–control studies with multiple disease states. *Biometrics* 2004; **60**:41–49.
10. Agresti A. *Categorical Data Analysis* (2nd edn). Wiley: New York, 2002.
11. Mccullagh P, Nelder JA. *Generalized Linear Models*. Chapman & Hall: London, 1983.
12. Cornfield J. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. *Federation Proceedings* 1962; **21**:58–61.
13. Mukherjee B, Ahn J, Liu I, Rathouz PJ, Sanchez BN. On elimination of nuisance parameters in a stratified proportional odds model by amalgamating conditional likelihoods. *Statistics in Medicine* 2008; **27**:4950–4971.
14. Lall R, Campbell MJ, Walters SJ, Morgan K. MRC CFAS Co-operative. A review of ordinal regression models applied on health-related quality of life assessments. *Statistical Methods in Medical Research* 2002; **11**:49–67.
15. Yee TW, Hastie TJ. Reduced-rank vector generalized linear models. *Statistical Modeling* 2003; **3**:15–41.
16. McCullagh P. Discussion of professor Anderson's paper. *Journal of Royal Statistical Society*, *Series B* 1984; **46**:1–30.
17. Haberman SJ. Tests for independence in two-way contingency tables based on canonical correlation and linear-by-linear interaction. *The Annals of Statistics* 1981; **9**:1178–1186.
18. Hanfelt JJ, Liang KY. Approximate likelihood ratios for general estimating functions. *Biometrika* 1995; **82**: 461–477.
19. Chen H, Chen J, Kalbfleisch JD. A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of Royal Statistical Society*, *Series B* 2001; **63**:19–29.
20. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall: London, 2000.
21. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Chapman & Hall: London, 2004.
22. 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.
23. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984; **6**:721–741.
24. Gelfand AE, Smith AFM. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 1990; **85**:398–409.
25. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 1992; **41**:337–348.
26. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences (with Discussion). *Statistical Science* 1992; **7**:457–472.
27. Sarma AV, Schottenfield D. Prostate cancer incidence, mortality, and survival trends in the United States: 1981–2001. *Seminars in Urologic Oncology* 2002; **20**:3–9.
28. Cooney KA, Strawderman MS, Wojno KJ, Doerr KM, Taylor A, Alcser KH, Heeringa SG, Taylor JM, Wei JT, Montie JE, Schottenfield D. Age-specific distribution of serum prostate-specific antigen in a community-based study of African-American men. *Urology* 2001; **57**:91–96.
29. Jewett HJ. The present status of radical prostatectomy for stages A and B prostatic cancer. *Urologic Clinics of North America* 1975; **2**:105–124.
30. Prostate IN. *American Joint Committee on Cancer*: *AJCC Cancer Staging Manual* (6th edn). Springer: New York, NY, 2002; 309–316.
31. Montie JE. Staging of prostate cancer: current TNM classification and future prospects for prognostic factors. *Cancer* 1995; **75**:1814–1818.