Scientific Research and Essays Vol. 5(11), pp. 1264-1273, 4 June, 2010
Available online at http://www.academicjournals.org/SRE
ISSN 1992-2248 ©2010 Academic Journals

Full Length Research Paper

A Bayesian analysis of bivariate ordered categorical responses using a latent variable regression model: Application to diabetic retinopathy data

Anoshirvan Kazemnejad[1], Farid Zayeri[2], Nor Aishah Hamzah[3], Rasool Gharaaghaji[4], Masoud Salehi[5]

[1]Department of Biostatistics, School of Medical Sciences, Tarbiat Modares University, P.O. Box: 14115-111, Tehran, Iran.
[2]Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
[3]Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia.
[4]Department of Biostatistics and Epidemiology, Faculty of Medicine, Urmia University of Medical Sciences, Urmia, Iran.
[5]Department of Statistics and Mathematics, Faculty of Management and Medical Informatics, Iran University of Medical Sciences, Tehran, Iran.

Accepted 25 March, 2010

Latent variable distribution models are frequently utilized for analyzing bivariate ordered categorical response data. In this context, choosing the bivariate normal distribution as the underlying latent distribution, which leads to the bivariate cumulative probit model, is the most common strategy for analyzing theses data sets. However, when the conditional distribution of the available bivariate response has an asymmetric form, other convenient asymmetric bivariate distributions may lead to a better fit. In this paper, we use an asymmetric bivariate cumulative latent variable distribution model for analyzing bivariate ordered categorical response data. For estimating the model parameters, we use two strategies: maximum likelihood and Bayesian approaches. We also use the proposed model for analyzing the data from 623 diabetic patients to identify some of the most important risk indicators of diabetic retinopathy among them. The obtained results revealed that patients' age at diagnosis, duration of diabetes, HbA1c, method of diabetes control, macular edema, and presence of hypertension and renal disease are significantly associated with the severity of diabetic retinopathy. In conclusion, both the maximum likelihood and Bayesian analyses resulted in similar significant risk indicators. However, it seems that the Bayesian analysis gives us smaller standard errors compared to the maximum likelihood approach.

Key words: Latent variable, bivariate ordinal response, asymmetric distribution, maximum likelihood estimation, Bayesian estimation, diabetic retinopathy.

INTRODUCTION

Diabetes is a major threat to global public health that is rapidly getting worse, and the biggest impact is on adults of working age in developing countries. The prevalence of diabetes is increasing worldwide, and most people will die or be disabled as consequence of vascular complications (The Advance Collaborative Group, 2008). According to the World Health Organization (WHO) reports, 171 million people had different types of diabetes in the beginning of the new century. In addition, the number of people suffering from diabetes is expected to rise to 366 million by 2030, with the most significant increases in developing countries (Wild et al., 2004). In the years (1995 - 2000), the prevalence of diabetes mellitus in Islamic Republic of Iran was, respectively, 5.5 and 5.7%. In addition, the WHO forecasts a prevalence rate of 6.8%

*Corresponding author. E-mail: kazem_an@modares.ac.ir. Tel: 98-21-82883875. Fax: 98-21-88006544.

for diabetes in our country in year 2025 (World Health Organization, 2006). This means that about 5.1 million Iranian people will suffer from diabetes in that year (King et al., 1998).

Diabetes can affect sight by causing cataracts, glaucoma, and most importantly, damage to blood vessels inside the eye, a condition known as "diabetic retinopathy". Diabetic retinopathy (DR) is a complication of diabetes that is caused by changes in the blood vessels of the retina. When blood vessels in the retina are damaged, they may leak blood and grow fragile, brush-like branches and scar tissue. This can blur or distort the vision images that the retina sends to the brain. Visual loss in DR patients is generally associated with sequelae from ischemia-induced neovascularization, diabetic macular edema and ischemic macular changes. Diabetic retinopathy is classified into an early stage, called "non-proliferative diabetic retinopathy" (NPDR), and a more advanced stage, "proliferative diabetic retinopathy" (PDR). PDR is a manifestation of ischemia-induced neovascularization from diabetes. NPDR stage can be classified into mild, moderate, severe, and very severe. Also, PDR stage is usually described as early, high-risk or advanced. In NPDR stage, retinal microvascular changes are limited to the confines of the retina and do no extend beyond the internal limiting membrane. NPDR can affect visual function through two mechanisms; variable degrees of intra-retinal capillary closure, resulting in macular ischemia and, increased retinal vascular permeability, resulting in macular edema. In PDR patients, extraretinal fibrovascular proliferation extends beyond the internal limiting membrane and is present in varying stages of development. The new vessels evolve in 3 stages; 1. Fine new vessels with minimal fibrous tissue appear, 2. The new vessels increase in size and extend, with an increased fibrous component, 3. The new vessels regress, leaving residual fibrovascular proliferation along the posterior hyaloid (Regillo, 2005).

In western countries, DR is one of the leading causes of visual impairment and blindness, especially in the working age group (Lim et al., 2008). DR is also one of the most important causes of vision loss in Asia (de Fine Olivarius et al., 2001). Epidemiologic studies in different countries, age groups and types of diabetes have reported different rates for prevalence of DR among diabetic patients (Al-Maskari and El-Sadig, 2007; Herman et al., 1998; El Haddad and Saad, 1998; Moss et al., 1994; Wong et al., 2006; Knudsen et al., 2006; Williams et al., 2004; Klein et al., 1989). A study in Iran, for instance, showed that the overall prevalence of retinopathy in patients with newly diagnosed diabetes mellitus was 13.8% (Abdollahi et al., 2006).

Bivariate correlated responses occur frequently in medical studies related to paired organs, such as eyes, ears, kidneys, lungs and so on. When we wish to analyze these kinds of data sets, the correlation between bivariate responses should be considered. In the previous decades, many statistical articles have been focused on analyzing different kinds of bivariate correlated data, such as continuous, binary or ordinal responses. In this context, choosing convenient modeling approaches is the most interesting methodology for analyzing these responses and describing the relationship between response data and a host of explanatory variables.

In screening studies related to diabetic patients, the ophthalmologists usually use an ordinal scale to determine the severity of diabetic retinopathy for each eye of a diabetic person. Therefore, for each diabetic patient a bivariate correlated ordinal response is available. Many data analysts previously proposed a variety of statistical methods for modeling bivariate correlated categorical responses including ophthalmic and diabetic retinopathy data sets. For instance, Rosner (1984) proposed two statistical methods for modeling normally and binomially outcomes with application ophthalmic data. Gange et al. (1995) presented several statistical methods for the analysis of ordered ophthalmic outcomes. Utilizing global odds ratio as a measure of association was suggested for analyzing bivariate ordered responses by Williamson et al. (1995). Another interesting approach for modeling bivariate ordered categorical data was suggested by Kim (1995).

In this paper, Kim proposed a bivariate cumulative probit regression model which uses stochastic ordering implicit in the data and the correlation coefficient of the bivariate normal distribution in expressing intra-subject dependency. He also used the data from the well known Wisconsin Epidemiologic Study of Diabetic Retinopathy to illustrate his approach. Biswas and Das (2002) developed a Bayesian approach for fitting the Kim's model. Zayeri and Kazemnejad (2006) introduced another bivariate cumulative regression model which may be useful when the bivariate latent variable distribution of the response data has an asymmetric form. They applied the proposed model for analyzing a bivariate correlated data from a periodontal study. In their model, a generalization of Gumbel's bivariate logistic distribution (Gumbel, 1961) which proposed by Satterthwaite and Hutchinson (1978) was used as the asymmetric latent variable instead of the bivariate normal distribution. In this paper, the model parameters were estimated using classic maximum likelihood (ML) approach.

In this manuscript, we use the data from the Tehran Epidemiologic Study of Diabetic Retinopathy to identify some of the most important effective factors on severity DR. Our main aim is to use both the maximum likelihood and Bayesian estimation approaches to fit the asymmetric bivariate cumulative regression model with generalization of Gumbel's bivariate logistic distribution as the latent variable. The utilized model and maximum likelihood estimation approach is entirely similar to those proposed by Zayeri and Kazemnejad (2006). The Bayesian approach for estimating the model parameters is the novel part of the present manuscript.

METHODS

Tehran epidemiologic study of diabetic retinopathy (TESDR)

The TESDR was an ophthalmic survey among diabetic patients in Tehran province performed by Ophthalmic Research Center of Shahid Beheshti University of Medical Sciences. The main objectives of this epidemiologic survey were:

(1) Determining raw and age-sex adjusted prevalence rates of different grades of DR among diabetic patients in this province.
(2) Identifying some of the most important risk factors or risk indicators of DR.

The study framework was consisted of a cohort of 639 diabetic patients who previously screened in another epidemiologic study of diabetes mellitus (carried out by the Center for Disease Control of the Ministry of Health) among 7500 inhabitants of Tehran province. To gather the basic information, these diabetic patients underwent physical and biochemical examinations.
Then, all patients were referred to a fellowship of retina in Negah Eye Clinic and underwent a complete ophthalmic examination. The presence of any grade of DR was recorded according to the new disease severity scale given by the American Academy of Ophthalmology (Wilkinson et al., 2003). More details about the TESDR including study population, sampling technique, methods of data collection, description of the data and preliminary analyses results was submitted as an epidemiologic-ophthalmic manuscript elsewhere (Javadi et al., 2009).

In this study, since the severity of DR was the most interested outcome for each eye of all examined patients, a correlated bivariate ordered response variable was available for each person (cluster). To reach the second objective of TESDR, we introduce convenient modeling approach and parameters estimation methods for describing the relationship between the explained bivariate response and a host of potential risk indicators of severity of DR.

Notation

Suppose an ophthalmic study with N participants and let y_{i1} and y_{i2} $(i = 1, \ldots, n)$ denote the bivariate ordered categorical outcome, respectively, corresponding to the right and left eyes of ith person, so, the bivariate response for person i can be denoted by $y_i = (y_{i1}, y_{i2})'$. Suppose, also, y_{i1} and y_{i2} can take one of the values $1, \ldots, K$. We assume that, they exist as non-observed continuous latent variables y_{i1}^* and y_{i2}^*, so that one can observe the outcome $y_{ij} = h$ if $\theta_{h-1} < y_{ij}^* \leq \theta_h$ for $j = 1{,}2$ with $\theta_0 = -\infty$ and $\theta_K = +\infty$. As usual, the θ_h's are called model cutoff points (or intercepts). In addition, suppose that p person-specific or eye-specific covariates are registered for each eye of all participants, therefore, for jth eye of ith person a $p \times 1$ covariate vector, say X_{ij}, is available. Assuming this, a $2 \times p$ design (covariate) matrix X_i for person i can be written as:

$$X_i = \begin{bmatrix} x'_{i1} \\ x'_{i2} \end{bmatrix}$$

Asymmetric bivariate latent distribution

Satterthwaite and Hutchinson introduced a generalization of Gumbel's bivariate distribution with the followings p.d.f. and c.d.f.:

$$f_v(x,y) = \frac{v(v+1)e^{-x}e^{-y}}{(1+e^{-x}+e^{-y})^{v+2}}$$

$$F_v(x,y) = (1+e^{-x}+e^{-y})^{-v}$$

Where, $v > 0$ is a parameter which describes the association between X and Y. They referred to v as "association parameter" (Satterthwaite and Hutchinson, 1978). In contrast with the bivariate normal distribution, the contours of the above mentioned bivariate distribution have an asymmetric form.

The correlation between variables X and Y can be computed using the generalized Riemann zeta function, that is $\zeta(s,a) = \sum_{m=0}^{\infty}(m+a)^{-s}$, and the association parameter v as

$$\rho = \frac{\zeta(2,v)}{\zeta(2,v) + \pi^2/6}$$

Simple calculations shows that if $v \longrightarrow 0$ then $\rho \longrightarrow 1$, and if $v \longrightarrow \infty$ then $\rho \longrightarrow 0$, thus in this distribution we have $0 < \rho < 1$.

Asymmetric latent variable model

For $h, l = 1{,}2, \ldots, K$, let the joint probability that ith person's right eye takes the value h and his/her left eye takes the value l is denoted by $\pi_{hl}(\omega; X_i)$, that is

$$\pi_{hl}(\omega; X_i) = pr(y_{i1} = h, y_{i2} = l \,|X_i)$$

Where, ω is $K + p$ dimensional vector of the unknown model parameters. In other words, ω includes $K - 1$ model cutoff points, θ's, p regression parameters, β's, and a correlation parameter, ρ . Note that, here, we assumed similar cutoff points and regression parameters for margins (right and left eyes). One can relax this assumption by treating different cutoff points and regression parameters by adding appropriate model parameters and modifying the design matrix. The interested reader can refer to Kim's paper for more details (Kim, 1995). Now, the cumulative joint probabilities can be shown as:

$$\gamma_{hl}(\omega; X_i) = pr(y_{i1} \leq h, y_{i2} \leq l \,|X_i) = \sum_{m=1}^{h} \sum_{n=1}^{l} \pi_{mn}$$

Assuming this, a family of bivariate latent distribution models with general link function G can be written as:

$$\gamma_{hl}(\omega; X_i) = G(\theta_h - x'_{i1}\beta, \theta_l - x'_{i2}\beta)$$

In this step, one can choose a convenient underlying bivariate distribution function as the link function G depending on the available data. For instance, if we choose the ordinary bivariate normal distribution as the link function, we have

$$\gamma_{hl}(\omega; X_i) = \varphi_\rho(\theta_h - x'_{i1}\beta, \theta_l - x'_{i2}\beta)$$

Where, φ_ρ is the standard bivariate normal distribution function with correlation coefficient ρ. Kim referred to this model as the bivariate cumulative probit model.

In this context, another choice for the link function G can be the bivariate distribution introduced by Satterthwaite and Hutchinson, F_υ. This choice leads to the following bivariate cumulative regression model:

$$\gamma_{hl}(\omega; X_i) = F_\upsilon(\theta_h - x'_{i1}\beta, \theta_l - x'_{i2}\beta)$$

Replacing F_υ in this model gives us the joint probabilities

$$\pi_{hl}(\omega; X_i) = \left[1 + e^{-(\theta_h - x'_{i1}\beta)} + e^{-(\theta_l - x'_{i2}\beta)}\right]^{-\upsilon} - \left[1 + e^{-(\theta_h - x'_{i1}\beta)} + e^{-(\theta_{l-1} - x'_{i2}\beta)}\right]^{-\upsilon}$$
$$- \left[1 + e^{-(\theta_{h-1} - x'_{i1}\beta)} + e^{-(\theta_l - x'_{i2}\beta)}\right]^{-\upsilon} + \left[1 + e^{-(\theta_{h-1} - x'_{i1}\beta)} + e^{-(\theta_{l-1} - x'_{i2}\beta)}\right]^{-\upsilon}$$

Because of asymmetric nature of the bivariate distribution suggested by Satterthwaite and Hutchinson, we refer to this model as the asymmetric bivariate cumulative model. Note that, in this model we used identical regression coefficients, β, and model cutoff points, θ, for both the left and right eyes of the units. For this model, the likelihood function can be written as

$$l(\omega) = \sum_{i=1}^{n}\sum_{h=1}^{K}\sum_{l=1}^{K} I_{i1h} I_{i2l} \log \pi_{hl}(\omega; X_i)$$

Kim (1995); Zayeri and Kazemnejad (2006) used a maximum likelihood estimation strategy for fitting the above mentioned models and estimating the parameters. In addition, Biswas and Das (2002) estimated the Kim's model parameters using a Bayesian approach. In the next section, we present a Bayesian method for fitting the described asymmetric bivariate cumulative model.

Bayesian analysis

Again, let (y_{i1}, y_{i2}) denotes the observed response vector for ith person under study, where y_{i1} and y_{i2} can take a value between 1 and K. We now suppose that, non-observed continuous latent variable $Y_i^* = (y_{i1}^*, y_{i2}^*)'$ follows the model

$$Y_i^* = X_i\beta + \varepsilon_i$$

Where, $\beta = (\beta_1, \ldots, \beta_p)'$ and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2})'$. We also postulate that ε_i follows the Satterthwaite and Hutchinson's bivariate distribution. The observed response vector can be denoted by $Y = (Y_1', \ldots, Y_i', \ldots, Y_n')'$ where, $Y_i = (y_{i1}, y_{i2})'$, and similarly, the latent variable vector is $Y^* = (Y_1^{*\prime}, \ldots, Y_i^{*\prime}, \ldots, Y_n^{*\prime})'$ with $Y_i^* = (y_{i1}^*, y_{i2}^*)'$. Assuming the same cutoff points and regression parameters for right and left eyes, the joint distribution of Y and Y^* is

$$f(Y, Y^* \mid \theta, \beta, \upsilon)$$
$$= \prod_{i=1}^{n}\left[\sum_{h,l \in S} I_{hl}^i(\theta_{h-1} < y_{i1}^* \le \vartheta_h, \theta_{l-1} < y_{i2}^* \le \theta_l)\right]$$
$$\times \frac{\upsilon(\upsilon+1)\exp[-(y_{i1}^* - x'_{i1}\beta)]\exp[-(y_{i2}^* - x'_{i2}\beta)]}{\{1 + \exp[-(y_{i1}^* - x'_{i1}\beta)] + \exp[-(y_{i2}^* - x'_{i2}\beta)]\}^{\upsilon+2}}$$

Where, $S = \{(h, l): h, l = 1, \ldots, K\}$ and

$$I_{hl}^i = \begin{cases} 1 & ; \quad y_{i1} = h \text{ and } y_{i2} = l \\ 0 & ; \quad O.W. \end{cases}$$

Then, the joint posterior distribution of the parameters can be written as

$$\pi(y^*, \theta, \beta, \upsilon | y) = \pi(y; y^* | \theta, \beta, \upsilon) \times \pi(\theta, \beta, \upsilon) \propto I(y; y^*, \theta) \times \pi(y^* | \beta, \upsilon) \times \pi(\theta, \beta, \upsilon)$$
$$= I(y; y^*, \theta) \times \pi(y^* | \beta, \upsilon) \times \pi(\theta) \times \pi(\beta) \times \pi(\upsilon); \; -\infty < y^*, \theta, \beta < +\infty, \upsilon > 0$$

To perform a Bayesian analysis, at first we postulate a multivariate normal distribution, that is, $N_p(\mu, \Sigma_0)$, for the prior distribution of β, a non-informative prior for θ and a Gamma distribution with parameters (α, λ) for υ. Assuming these priors, the following conditional posterior distributions are obtained using Markov Chain Monte Carlo (MCMC) sampling technique

$$\pi\left(y_i^* \mid \theta, \beta, \upsilon, y_i = \binom{h}{l}\right)$$
$$\propto \left[\sum_{h,l \in S} I_{hl}^i(\theta_{h-1} < y_{i1}^* \le \theta_h, \theta_{l-1} < y_{i2}^* \le \theta_l)\right]$$
$$\times \frac{\upsilon(\upsilon+1)\exp[-(y_{i1}^* - x'_{i1}\beta)]\exp[-(y_{i2}^* - x'_{i2}\beta)]}{\{1 + \exp[-(y_{i1}^* - x'_{i1}\beta)] + \exp[-(y_{i2}^* - x'_{i2}\beta)]\}^{\upsilon+2}}$$
$$\propto I\begin{bmatrix}\theta_{h-1} < y_{i1}^* \le \theta_h \\ \theta_{l-1} < y_{i2}^* \le \theta_l\end{bmatrix} \times \frac{\exp[-(y_{i1}^* - y_{i2}^*)]}{\{1 + \exp[-(y_{i1}^* - x'_{i1}\beta)] + \exp[-(y_{i2}^* - x'_{i2}\beta)]\}^{\upsilon+2}}$$

This is a truncated Satterthwaite and Hutchinson's bivariate distribution over the above domain. Here, the conditional posterior distribution of θ given the others is

$$\pi\left(\binom{\theta_h}{\theta_l} \mid \theta_{-\binom{h}{l}}, \beta, v, y^*, y\right)$$

$$\propto \prod_{i=1}^{n}\left\{ I_{h,l}^{i}\begin{bmatrix}\theta_{h-1} < y_{i1}^* \le \theta_h \\ \theta_{l\;1} < y_{i2}^* < \theta_l\end{bmatrix} + I_{h+1,l}^{i}\begin{bmatrix}\theta_h < y_{i1}^* \le \theta_{h+1} \\ \theta_{l\;1} < y_{i2}^* < \theta_l\end{bmatrix}\right.$$

$$\left. + I_{h,l+1}^{i}\begin{bmatrix}\theta_{h-1} < y_{i1}^* \le \theta_h \\ \theta_l < y_{i2}^* \le \theta_{l+1}\end{bmatrix} + I_{h+1,l+1}^{i}\begin{bmatrix}\theta_h < y_{i1}^* \le \theta_{h+1} \\ \theta_l < y_{i2}^* \le \theta_{l+1}\end{bmatrix}\right\}$$

Where, $\theta_{-\binom{h}{l}}$ is the collection of θ_u and θ_t except θ_h and θ_l.

This conditional distribution can be considered to be uniform over the same regions defined by Biswas and Das (2002); Albert and Chib (1993) that is

$$\theta_{cj} \sim unif\,(t_{c\theta}, r_{c\theta})$$

$$t_{c\theta} = max\left\{\max_{i=1,2,\dots,n}\{y_{ij}^* : y_{ij} = c\}, \theta_{c-1,j}\right\}$$

$$r_{c\theta} = min\left\{\min_{i=1,2,\dots,n}\{y_{ij}^* : y_{ij} = c+1\}, \theta_{c+1,j}\right\}$$

Where, $y_{ij} = c \;\; if\ and\ only\ if \;\; \theta_{c-1,j} < y_{ij}^* < \theta_{cj}$, for $j = 1, 2$.

Assuming the above mentioned multivariate normal distribution for β, the conditional distribution of this parameter is

$$\pi(\beta \mid y^*, \theta, v, y)$$

$$\propto \prod_{i=1}^{n} \frac{v(v+1)exp[-(y_{i1}^* - x_{i1}'\beta)]exp[-(y_{i2}^* - x_{i2}'\beta)]}{\{1 + exp[-(y_{i1}^* - x_{i1}'\beta)] + exp[-(y_{i2}^* - x_{i2}'\beta)]\}^{v+2}}$$

$$\times \left(\frac{1}{2\pi|\Sigma_0|}\right)^{p/2} exp[-1/2(\beta - \mu)'\Sigma_0^{-1}(\beta - \mu)]$$

$$\propto exp\left\{-\frac{1}{2(\beta - \mu)'\Sigma_0^{-1}(\beta - \mu)}\right.$$

$$+ \sum_{i=1}^{n}[x_{i1}'\beta + x_{i2}'\beta$$

$$\left. - (v+2)ln\{1 + exp[-(y_{i1}^* - x_{i1}'\beta)] + exp[-(y_{i2}^* - x_{i2}'\beta)]\}]\right\}$$

which is an unknown distribution. Again, using Metropolis-Hastings algorithm with the distribution $N_p\left(\hat{\theta}, \left[I(\hat{\theta})^{-1}\right]\right)$, the following normal distribution is obtained after straightforward calculations for the Hessian matrix

$$N\left(\hat{\beta}, \left(\Sigma_0^{-1} + (v+2)\sum_{i-1}^{n} X_i X_i' \frac{exp[-(y_i^* - X_i'\beta)]}{(1 + exp[-(y_i^* - X_i'\beta)])^2}\right)^{-1}\right)$$

The vector $\hat{\beta}$ can be obtained using a Gauss-Newton numerical method.

In addition, postulating the described Gamma distribution for v, the conditional posterior distribution of this parameter can be written as

$$\pi(v \mid \theta, \beta, y^*, y) = \prod_{i=1}^{n} \frac{v(v+1)exp[-(y_{i1}^* - x_{i1}'\beta)]exp[-(y_{i2}^* - x_{i2}'\beta)]}{\{1 + exp[-(y_{i1}^* - x_{i1}'\beta)] + exp[-(y_{i2}^* - x_{i2}'\beta)]\}^{v+2}} \times \frac{\lambda^{\alpha} v^{\alpha-1} e^{-\lambda v}}{\Gamma(\alpha)}$$

$$\propto v^{n+\alpha-1}(v$$

$$+ 1)^n exp\left[-v\left(\lambda + \sum_{i=1}^{n} ln(1 + exp[-(y_{i1}^* - x_{i1}'\beta)] + exp[-(y_{i2}^* - x_{i2}'\beta)])\right)\right]$$

$$\propto (v+1)^n \Gamma\left[n + \alpha, \left[\lambda + \sum_{i=1}^{n} ln(1 + exp[-(y_{i1}^* - x_{i1}'\beta)] + exp[-(y_{i2}^* - x_{i2}'\beta)])\right]\right]$$

This conditional distribution has an unknown form too, thus, we can utilize a Metropolis-Hastings algorithm with the following distribution for generating the required random sample

$$\Gamma\left[n + \alpha, \left[\lambda + \sum_{i=1}^{n} ln(1 + exp[-(y_{i1}^* - x_{i1}'\beta)] + exp[-(y_{i2}^* - x_{i2}'\beta)])\right]\right]$$

In the final step, the Gibbs sampler was implemented using convenient initial guesses for θ, β, v and y^* for simulating the conditional distributions and obtaining a sample from $(y^*, \theta, \beta, v \mid y)$.

RESULTS

Generally, in Tehran Epidemiologic Study of Diabetic Retinopathy (TESDR), a total sample of 623 type 2 diabetic patients underwent ophthalmic examination to identify the status of DR in their eyes. As mentioned before, ophthalmologists usually record the severity of DR as an ordinal scale with rather complex categories. In this paper, we use a simplified form of this scale with the following categories: 1. normal, 2. mild NPDR, 3. moderate NPDR, 4. severe NPDR and 5. PDR. In addition, because of small sample size for patients with severe DR, we combined the categories moderate and severe NPDR in our analysis. Table 1 shows the joint distribution of severity of DR in the left and right eyes of this sample. As we can see, more than 90% of the data lies in the main diagonal of this 4 × 4 contingency table of the bivariate response. This means that there is a strong correlation between severity of DR in left and right eyes of the patients. A Spearman's coefficient correlation of 0.948 is a strong evidence for this significant correlation.

In the modeling process, we used the following explanatory variables to identify some of the most important risk indicators of DR; age at diagnosis (in years), sex (male, female), duration of diabetes (<5, 5 - 10, 10 - 15, 15 - 20 and >20 yrs), HbA1c (values less than 7% were considered as "controlled", values more than or equal to 7% considered as "uncontrolled"), method of diabetes control (insulin injection, else including exercise, diet and oral medication), renal disease (presence, absence) and macular edema (presence, absence). In this study, all the patients were

Table 1. Severity of diabetic retinopathy.

		Left eye				Total
		Normal	**Mild**	**Moderate- Severe**	**Proliferative**	
Right eye	Normal	399	11	0	0	410
	Mild	11	86	3	0	100
	Moderate- Severe	0	0	59	11	70
	Proliferative	0	0	6	37	43
	Total	410	97	68	48	623

Table 2. Characteristics of the diabetic patients under study.

Characteristics	Category	Mean ±SD	No	Percent
Age at diagnosis		51.06 ± 11.66		
Sex	Female		339	54.4
	Male		284	45.6
Duration of diabetes	<5		210	33.7
	5-10		212	34.0
	10-15		100	16.1
	15-20		45	7.2
	>20		56	9.0
HbA1c	Controlled		390	62.6
	Uncontrolled		233	37.4
Method of diabetes control	Insulin injection		58	9.3
	Else		565	90.7
Diabetic nephropathy	Absence		588	94.4
	Presence		35	5.6
Macular edema	Absence		586	94.1
	Presence		37	5.9

examined by an endocrinologist and the presence of diabetic nephropathy was based on the diagnosis of this physician. The definition of the diabetic nephropathy was considered as: “the presence of persistent proteinuria (albumin excretion rate > 200 µg/min or 300 mg/day) in the absence of other causes”. Table 2 shows the summary statistics for these explanatory variables.

In the next step, we fitted asymmetric bivariate cumulative model to the data using the ordinary maximum likelihood approach. Table 3 shows the obtained results. These findings tell us that the described explanatory variables had significant effect on severity of DR, except a category of diabetes duration (category 5 -10 years compared to more than 20 years duration of diabetes). As shown in Table 3, the ML estimate for association parameter ($\hat{\upsilon}$) is 0.312. Replacing this value in the generalized Riemann zeta functions and consequently in the illustrated equation for the correlation parameter results in an estimate of 0.874 for ρ.

We could also interpret the estimates in terms of latent variable scale. For instance, change from absence to presence of renal disease and macular edema lead to increase of 0.282 and 1.908 in the latent variable scale, respectively. Combining these effects gives us a total change of 2.190 in the latent variable scale. This means that simultaneous presence of renal disease and macular edema can lead to change in diabetic retinopathy severity from mild $\left(4.226 = \hat{\theta}_1 \leq y_i^* < \hat{\theta}_2 = 5.368\right)$ to the worst status, that is, proliferative $\left(6.238 = \hat{\theta}_3 \leq y_i^*\right)$.

In the final stage of the modeling process, we estimated the parameters of our proposed model using the

Table 3. Maximum likelihood estimates for the asymmetric bivariate cumulative model.

Parameter		Estimate	SE	95% CI	P-value
θ_1		4.226	0.692	[2.870-5.583]	<0.001
θ_2		5.368	0.698	[4.000-6.737]	<0.001
θ_3		6.238	0.699	[4.868-7.608]	<0.001
υ		0.312	0.082	--	--
Age at diagnosis		0.025	0.007	[0.011-0.039]	0.001
Sex	Male	0.247	0.114	[0.024-0.470]	0.030
	Female		Reference category		
Duration	>20	1.484	0.253	[0.989-1.980]	<0.001
	15-20	1.221	0.245	[0.734-1.698]	<0.001
	10-15	0.678	0.226	[0.236-1.121]	0.003
	5-10	0.308	0.194	[-0.072-0.689]	0.112
	<5		Reference category		
HbA1c	Uncontrolled	0.270	0.136	[0.003-0.537]	0.048
	Controlled		Reference category		
Method of diabetes control	Insulin injection	0.377	0.159	[0.065-0.689]	0.018
	Else		Reference category		
Diabetic nephropathy	Presence	0.282	0.115	[0.057-0.507]	0.014
	Absence		Reference category		
Macular edema	Presence	1.908	0.235	[1.447-2.369]	<0.001
	Absence		Reference category		

explained Bayesian analysis. Table 4 shows the obtained posterior summary statistics. As we can see, the resulted parameter estimates using Bayesian analysis are rather similar to those obtained by maximum likelihood approach. However, it seems that, the standard errors from Bayesian analysis are considerably smaller than ML estimates. In addition, we found a larger estimate for association parameter, $\hat{\upsilon} = 0.372$, and consequently smaller correlation parameter, $\hat{\rho} = 0.834$, in the Bayesian analysis.

DISCUSSION

In this paper, we used an asymmetric bivariate latent distribution model for analyzing correlated ordered categorical data. We applied the proposed model for analyzing diabetic retinopathy data to identify some of the risk indicators of this ophthalmic disease in Iranian population. However, the application of this model is not limited to ophthalmic data. Apparently, we can utilize this model for analyzing data from other studies related to paired organs such as ears, lungs, kidneys, and so on, where the available outcome is a highly correlated ordered categorical data.

In this study, we only utilized the person-specific covariates for modeling diabetic retinopathy data. As mentioned before, one can use both the person-specific and organ-specific (eye-specific) covariates for modeling these kinds of data sets. We could also extend the model by allowing identical regression parameters for the person-specific covariates and different regression parameters for the organ-specific covariates. To do this, we could write the $2 \times (p + q)$ design matrix as

$$X_i = \begin{bmatrix} x_i' & x_{i1}' \\ x_i' & x_{i2}' \end{bmatrix}$$

Then, the extended model can be written as

$$\gamma_{hl}(\omega; X_i) = F_\upsilon(\theta_h - x_i'\beta - x_{i1}'\delta_1, \theta_l - x_i'\beta - x_{i2}'\delta_2)$$

Where, β is now a $p \times 1$ vector of identical regression

Table 4. Bayesian estimates for the asymmetric bivariate cumulative model.

Parameter		Mean	S.E.	MC error	2.5%	Median	97.5%
θ_1		3.8040	0.1554	0.0232	3.5543	3.7900	4.1625
θ_2		4.7340	0.1806	0.0259	4.3723	4.7240	5.1390
θ_3		5.2460	0.3414	0.0275	4.0860	5.2250	5.6893
υ		0.3719	0.0589	0.0067	0.6832	0.7927	0.9117
Age at diagnosis		0.0300	0.0011	0.0000	0.0279	0.0299	0.0322
Sex	Male	0.3932	0.0590	0.0013	0.2791	0.3928	0.5071
	Female			Reference category			
Duration	>20	1.3010	0.1068	0.0024	1.0915	1.3000	1.5064
	15-20	1.2551	0.1134	0.0025	1.0331	1.2562	1.4824
	10-15	0.4843	0.1440	0.0031	0.2083	0.4838	0.7650
	5-10	0.3311	0.0601	0.0012	0.2110	0.3305	0.4503
	<5			Reference category			
HbA1c	Uncontrolled	0.3504	0.0890	0.0031	0.1736	0.3507	0.6206
	Controlled			Reference category			
Method of diabetes control	Insulin injection	0.5349	0.1082	0.0025	0.3228	0.5355	0.7442
	Else			Reference category			
Diabetic nephropathy	Presence	0.2656	0.0726	0.0015	0.1252	0.2652	0.4085
	Absence			Reference category			
Macular edema	Presence	1.8620	0.1444	0.0031	1.5784	1.8560	2.1502
	Absence			Reference category			

parameters for the person-specific covariates, and δ_1 and δ_2 are different $q \times 1$ vectors of regression coefficients for the eye-specific covariates.

In addition, one can extend the suggested model by allowing different cutoff points for the marginal distributions and different regression parameters for the margins. This needs the fallowing $2 \times (p + 2q)$ design matrix

$$X_i = \begin{bmatrix} x_i' & x_{i1}' & 0' \\ x_i' & 0' & x_{i2}' \end{bmatrix}$$

and the extended model

$$\gamma_{hl}(\omega; X_i) = F_i(\theta_{h1} - x_i'\beta_1 - x_{i1}'\delta_1, \theta_{l2} - x_i'\beta_2 - x_{i2}'\delta_2)$$

As said by Kim, the use of such model with different regression parameters for the margins heavily depends on the available bivariate response (1995). In diabetic retinopathy data, for instance, postulating different regression parameters for the margins may lead to the complexity of interpretation of the estimates, especially for the physicians, and gives us no additional information about the behaviour of the margins. However, one can use this extension for situations where subjects are classified using two different categories in the margins.

To obtain the maximum likelihood estimates of the model parameters, we used the function *nlminb* in the S-PLUS software. Using a straightforward program for defining the described likelihood function, we can obtain both the parameter estimates and their standard errors. For fitting the model using the Bayesian approach, an R program was utilized. As we can see in the methods section, the Bayesian approach is computationally more complex and time-consuming compared to the maximum likelihood strategy. Here, one may ask this question 'why we should estimate the model parameters using the Bayesian method?' When a data with small sample size is available, the maximum likelihood strategy may lead to larger standard errors and unrealistic non-significant estimates. In these situations, the Bayesian approach (which may results in smaller standard errors and more sensible significant estimates) could be the preferable estimating method. In the present study, however, both the ML and Bayesian approaches led to similar significant parameters, except for category 5 - 10 years of duration of diabetes. This is probably due to large sample size data in our study. On the other hand, the Bayesian approach resulted in smaller standard errors than the ML method, even in this large sample size data.

In Bayesian analysis, to perform the required computations by Gibbs sampler, we initially employed 5000 updates burn-in followed by additional 5000 updates which used for obtaining the posterior summary statistics. We also used the Gelman and Rubin's approach to ensure the convergence of the Gibbs sampler. In this context, choosing convenient initial guesses for parameter estimates plays an important role in decreasing the required time for the convergence. We used the obtained estimates from the generalized estimating equations (GEE) approach as the initial values of the model parameters. Starting from these initial guesses, we generated a sample size of 10000. Then, to minimize the effect of initial values, 5000 replications were deleted as the burning samples and the remaining 5000 replications were used to approximate the posterior distribution.

In modeling process, because of inadequate sample size for some categories of explanatory variables, we combined the small sample size categories with the others. In data collecting stage, for instance, we gathered information about all methods of diabetes control (including: diet and/or exercise, oral medication, and insulin injection). In the data analysis stage, since the sample size for diet and/or exercise group was inadequate, we combined this category with oral medication group to estimate the model parameters properly. Small sample size for some categories of explanatory variables was one of the limitations of our study. In addition, we could not collect reliable information about different drugs (for instance, oral medication for diabetes control or anti-hypertensive drugs) used by the patients under study, because of recall bias and wide variety of these medications. This may be considered as the other limitation of our research.

Conclusions

In this paper, we utilized two estimating methods, ML and Bayesian approaches, for estimating the parameters of a bivariate cumulative model. This model is a convenient choice for analyzing bivariate ordered categorical data with an asymmetric underlying latent distribution. Analyzing the diabetic retinopathy data from TESDR study showed that both the estimating methods gives us rather similar parameter estimates, but the Bayesian approach leads to smaller standard errors. Therefore, using the Bayesian analysis is recommended, especially when a small sample size data set is available.

ACKNOWLEDGEMENTS

Hereby, we wish to thank all of our colleagues in Labbafinejad Ophthalmic Research Center, especially Dr Nasrin Rafati, Dr Mehdi Yaseri, Dr Marzieh Katibeh and Dr Hamid Ahmadieh for providing the TESDR data.

REFERENCES

Abdollahi A, Malekmadani MH, Mansoori MR, Bostak A, Abbaszadeh MR, Mirshahi A (2006). Prevalence of diabetic retinopathy in patients with newly diagnosed type II diabetes mellitus. Acta. Medica. Iranica, 44: 415-419.

Albert J, Chib S (1993). Bayesian analysis of binary and polychotomous response data. J. Am. Stat. Assoc. 88:669-679.

Al-Maskari F, El-Sadig M (2007). Prevalence of diabetic retinopathy in the United Arab Emirates: a cross-sectional survey. BMC. Ophthalmol. pp. 7-11.

Biswas A, Das K (2002). A Bayesian analysis of bivariate ordinal data: Wisconsin epidemiologic study of diabetic retinopathy revisited. Stat. Med., 21: 549-559.

de Fine Olivarius N, Nielsen NV, Andreasen AH (2001). Diabetic retinopathy in newly diagnosed middle-aged and elderly diabetic patients. Prevalence and interrelationship with microalbuminuria and triglycerides. Graefes. Arch. Clin. Exp. Ophthalmol., 239: 664-672.

El Haddad OA, Saad MK (1998). Prevalence and risk factors for diabetic retinopathy among Omani diabetics. Br. J. Ophthalmol., 82: 901-906.

Gange SJ, Linton KLP, Scott AJ, Demets DL, Klein R (1995). A comparison of methods for correlated ordinal measures with ophthalmic applications. Stat. Med., 14: 1961-1974.

Gumbel EJ (1961). Bivariate logistic distribution. J. Am. Stat. Assoc., 56: 335-349.

Herman WH, Aubert RE, Engelgau MM, Thompson TJ, Ali MA, Sous ES (1998). Diabetes mellitus in Egypt: glycaemic control and microvascular and neuropathic complications. Diabet. Med., 15: 1045-1051.

Javadi MA, Katibeh M, Rafati N, Dehghan MH, Zayeri F, Yaseri M, Sehat M, Ahmadieh H (2009). Prevalence of diabetic retinopathy in Tehran province: a population-based study. BMC. Ophthalmol. pp. 9-12.

Kim K (1995). A bivariate cumulative probit regression model for ordered categorical data. Stat. Med., 14: 1341-1352.

King H, Aubert RE, Herman WH (1998). Global burden of diabetes, 1995-2025. Diabetes. Care, 21: 1414-1431.

Klein R, Klein BEK, Moss SE (1989). The Wisconsin epidemiological study of diabetic retinopathy: a review. Diabetes. Metab. Rev., 5: 559-570.

Knudsen LL, Lervang HH, Lundbye-Christensen S, Gorst-Rasmussen A (2006). The North Jutland County Diabetic Retinopathy Study: population characteristics. Br. J. Ophthalmol., 90: 1404-1409.

Lim MCC, Lee SY, Cheng BCL, Wong DWK, Ong SG, Ang CL (2008). Diabetic Retinopathy in Diabetics Referred to a Tertiary Centre from a Nationwide Screening Programme. Ann. Acad. Med. Singapore, 37: 753-759.

Moss SE, Klein R, Klein BE (1994). Ten-year incidence of visual loss in a diabetic population. Ophthalmology, 101: 1061-1070.

Regillo C (2005). Basic and clinical science course. Section 12. Retina and vitreous. San Francisco: American Academy of Ophthalmology; pp. 99-119.

Rosner B (1984). Multivariate methods in ophthalmology with application to other paired-data situations. Biometrics. 40: 1025-1033.

Satterthwaite SP, Hutchinson TP (1978). A generalization of Gumbel's bivariate logistic distribution. Metrika, 25: 163-170.

Sundling V, Gulbrandsen P, Jervell J, Straand J (2008). Care of vision and ocular health in diabetic members of a national diabetes organization: a cross-sectional study. BMC. Health. Serv. Res., pp. 8-159.

Wild S, Roglic G, Green A, Sicree R, King H (2004). Global Prevalence of Diabetes: Estimates for the year 2000 and projections for 2030. Diabetes. Care, 27: 1047-1053.

Wilkinson CP, Ferris FL, Klein RE (2003). Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology, 110: 1677-1682.

Williams R, Airey M, Baxter H, Forrester J, Kennedy-Martin T, Girach A

(2004). Epidemiology of diabetic retinopathy and macular oedema: a systematic review. Eye. 18: 963-983.

Williamson JM, Kim K, Lipsitz SR (1995). Analyzing bivariate ordinal data using a global odds ratio. J. Am. Stat. Assoc. 90: 1432-1437.

Wong TY, Klein R, Islam FM, Cotch MF, Folsom AR, Klein BE (2006). Diabetic retinopathy in a multi-ethnic cohort in the United States. Am. J. Ophthalmol., 141: 446-455.

World Health Organization (2006). Guidelines for the prevention, management and care of diabetes mellitus. EMRO Technical publications series 32, Geneva.

Zayeri F, Kazemnejad A (2006). A Latent Variable Regression Model for symmetric Bivariate Ordered Categorical Data. J. Appl. Stat., 33: 745-755.