Bivariate Versus Univariate Ordinal Categorical Data with Reference to an Ophthalmologic Study

Jean-François Angers* Atanu Biswas†

CRM-2858

July 2002

*Département de mathématiques et de statistique, Université de Montréal C.P. 6128, Succ. "Centre-ville", Montréal, Québec, H3C 3J7 angers@dms.umontreal.ca

†Applied Statistics Unit, Indian Statistical Institute 203 B.T. Road, Calcutta—700 035, India atanu@isical.ac.in

Abstract

The Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) is a population-based epidemiologic study carried out in Southern Wisconsin during the eighties of the last century. The resulting data were analyzed by different statisticians and ophthalmologists during the last two decades. Most of the analyses were carried out on the baseline data, although there were two follow-up studies on the same population. A Bayesian analysis of the first follow-up data, taken four years after the baseline study, was carried out by Angers and Biswas (2001) where the choice of the best model in terms of the covariate inclusion is done and estimates of the associated covariate effects were obtained using the baseline data to set the prior for the parameters. In the present article we consider an univariate transformation of the bivariate ordinal data, and a parallel analysis with the much simpler univariate data is carried out. The results are then compared with the results of Angers and Biswas (2001). It is concluded that the transformation to univariate data is suitable in the present context as the analysis of the transformed data catches most of the features of the bivariate data.

Key words: Univariate ordinal data, bivariate ordinal data, sensitivity analysis, Bayesian selection model.

1 Introduction and Data Description

Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) is a population-based study in Southern Wisconsin between 1980 and 1982, in which a total of 996 insulin-taking, younger onset diabetic persons were examined using standard protocols to determine the prevalence and severity of diabetic retinopathy and associated risk variables (see Klein, Klein, Moss, Davis and DeMets, 1984a, b). The basic goal of the study (*cf.* Klein *et al.*, 1984a, b) was to find the associated risk factors which are important in planning a well-coordinated approach to the public health problem posed by the complications of diabetes. There were 4-year and 10-year follow-up examinations (*cf.* Klein, Klein, Moss, Davis and DeMets, 1989, and Klein, Klein, Moss and Cruickshanks, 1994).

The observations corresponding to any individual (at any time point) was a bivariate ordinal categorical in nature (in the presence of several covariates). Both the right and left eye retinopathy severity levels are recorded as two components of the bivariate response. Possible values are 10, 21, 31, 37, 43, 47, 53, 60, 61, 65, 71, 75, 85 corresponding to increasing levels of severity of retinopathy within an eye.

Three eye-specific covariates are recorded. These are right and left eye macular edema (ME) (present/absent), right and left eye refractive error (RE) in diopters, right and left eye intraocular pressure (IOP) in mmHg. In addition 8 person-specific covariates are recorded, namely the duration of diabetes (DuD) in years, glycosylated hemoglobin (GH) in percent, systolic and diastolic blood pressures (SBP & DBP) are measured in mmHg., body mass index (BMI) in kilograms per meter squared, pulse rate (PR) in beats per 30 seconds, urine protein (UP) (present/absent) and doses of insulin (DI) per day.

Analysis of ordinal categorical data becomes much more complicated when the ordinal categorical data is of multivariate in nature. Dale's (1986) paper opened the floodgate followed by a deluge of papers in this direction. Many of the early analyses were in the frequentist's set up (see e.g. Williamson, Kim and Lipsitz, 1995; Molenberghs and Lessafre, 1994; Kim, 1995; Williamson and Kim, 1996; Kim, Lipsitz and Williamson, 1996; Williamson, Lipsitz and Kim, 1999). Note that, most of the frequentist's approaches are computationally quite expensive.

In a Bayesian paradigm, Biswas and Das (2002) considered a model using normally distributed latent variables similar to that of Kim (1995), where one may easily arrive at a consistent solution of the underlying regression parameter and may draw inference through the well known Gibbs sampler approach. Note that all the above mentioned works were done using the baseline data only. Das and Biswas (2001) carried out a Bayesian semiparametric analysis using both the baseline and 4-year follow-up data.

Angers and Biswas (2001) carried out a Bayesian analysis of the 4-year follow-up data using the baseline data to fix the priors for different parameters. In the present paper, we first provide a brief description of the methodology and the results of Angers and Biswas (2001), in Section 2. In Section 3, we discuss a transformation of the bivariate data to an univariate one, and provide the methodology of analyzing such univariate data. Quite naturally, the technique of analysis will be simpler and the computational task will be easier. The results with the conclusions are presented in Section 4.

2 Analysis of Bivariate Data

Let y_{Li} and y_{Ri} denote the bivariate ordered categorical responses for the ith individual corresponding to left and right eye respectively. Note that, $y_{Li}, y_{Ri} \in \{10, 21, 31, 37, 43, 47, 53, 60, 61, 65, 71, 75, 85\}$. Let y_L and y_R be the vectors combining y_{Li}'s and y_{Ri}'s for all the individuals. In order to assume normality of the error terms, we add z_{Li} and z_{Ri} to y_{Li} and y_{Ri}, where $(z_{Li}, z_{Ri})^T \sim N_2(0, \sigma_Z^2 \mathrm{I}_\rho)$ with I_ρ is the 2×2 correlation matrix with unknown correlation ρ, and σ_z^2 is such that $y_{Li} \pm 3\sigma_z$ and $y_{Ri} \pm 3\sigma_z$ will not change categories. Let z_L (z_R) be the vector combining all the z_{Li}'s (z_{Ri}'s).

Let $\mathrm{u}_L = \mathrm{y}_L + \mathrm{z}_L$ and $\mathrm{u}_R = \mathrm{y}_R + \mathrm{z}_R$ and these u_L and u_R are the true values and we observe y_L and y_R in place of them. We model u_L and u_R as follows:

$$\mathrm{u} = \mathrm{X}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

where

$$\mathrm{u} = \begin{pmatrix} \mathrm{u}_L \\ \mathrm{u}_R \end{pmatrix}, \mathrm{X} = \begin{pmatrix} \mathrm{X}_0 & \mathrm{X}_1 & 0 \\ \mathrm{X}_0 & 0 & \mathrm{X}_2 \end{pmatrix}, \boldsymbol{\theta} = (\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T, \boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_L^T, \boldsymbol{\epsilon}_R^T)^T,$$

$\boldsymbol{\beta}_0$ are the covariate common to both eyes, $\boldsymbol{\beta}_1$, those specific for the left eye only and $\boldsymbol{\beta}_2$, those for the right eye only and $\boldsymbol{\epsilon} \sim N_{2n}(0, \sigma^2 \mathrm{I}_{2n})$.

For the p-component parameter vector $\boldsymbol{\theta}$, we consider

$$\boldsymbol{\theta} \sim N_p\left(\boldsymbol{\theta}_0, \frac{\sigma^2}{\kappa}\mathrm{V}\right), \tag{1}$$

where the hyperparameters $\boldsymbol{\theta}_0$ and κ ($\kappa \leq 1$) are assumed to be known and σ^2 is unknown. An inverted gamma prior with parameters α and β is considered for σ^2. The prior of ρ is chosen to be an uniform density on the interval $(-1, 1)$. We use the baseline data to estimate $\boldsymbol{\theta}_0$. To denote the baseline data we just put "*" to y_L, y_R and X. Thus $\boldsymbol{\theta}_0$ is estimated using y^*_L, y^*_R and X^* as follows:

$$\widehat{\boldsymbol{\theta}}_0 = (\mathrm{X}^{*T}\mathrm{X}^*)^{-1}\mathrm{X}^{*T}\mathrm{y}^*,$$

where $\mathrm{y}^* = (\mathrm{y}^{*T}_L, \mathrm{y}^{*T}_R)^T$. Sensitivity analysis on κ is also needed to be done.

Using standard technique, after some routine steps, it can be shown that the conditional posteriors of $\boldsymbol{\theta}$, σ^2 and ρ are

$$\begin{aligned}
\boldsymbol{\theta}|\mathrm{u}, \sigma^2, \rho &\sim N_p\big((\kappa\mathrm{V}^{-1} + \mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1}(\kappa\mathrm{V}^{-1}\boldsymbol{\theta}_0 + \mathrm{X}^T\mathrm{U}^{-1}\mathrm{X}\boldsymbol{\theta}_{LS}(\rho)), \\
&\qquad\qquad \sigma^2(\kappa\mathrm{V}^{-1} + \mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1}\big), \\
\sigma^2|\rho, \mathrm{u} &\sim I\Gamma\bigg(\frac{\alpha + n - 2}{2}, 0.5 \times \big[\beta + \mathrm{r}^T\mathrm{U}^{-1}\mathrm{r} + (\boldsymbol{\theta}_{LS}(\rho) - \boldsymbol{\theta}_0)^T \\
&\qquad\qquad \times [\kappa^{-1}\mathrm{V} + (\mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1}]^{-1}(\boldsymbol{\theta}_{LS}(\rho) - \boldsymbol{\theta}_0)\big]\bigg), \\
\pi_3(\rho|\mathrm{u}) &\propto \frac{1}{|\mathrm{D} - \rho\mathrm{E} + (1-\rho^2)\kappa\mathrm{V}^{-1}|^{1/2}} \\
&\qquad\qquad \times \left(\frac{(1-\rho^2)}{\mathrm{r}^T\mathrm{r} - 2\rho t + (1-\rho^2)[\beta + h(\rho)]}\right)^{(\alpha+n)/2},
\end{aligned}$$

where

$$\begin{aligned}
& r = \mathrm{u} - \mathrm{X}\boldsymbol{\theta}_{LS}(\rho) = (\mathrm{r}^T_L, \mathrm{r}^T_R)^T, \qquad \boldsymbol{\theta}_{LS}(\rho) = (\mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1}\mathrm{X}^T\mathrm{U}^{-1}\mathrm{u}, \\
& \mathrm{U} = \begin{pmatrix} \mathrm{I} & \rho\mathrm{I} \\ \rho\mathrm{I} & \mathrm{I} \end{pmatrix}, \qquad\qquad t = \mathrm{r}^T_L\mathrm{r}_R, \\
& \mathrm{D} = 2\mathrm{X}^T_0\mathrm{X}_0 + \mathrm{X}^T_1\mathrm{X}_1 + \mathrm{X}^T_2\mathrm{X}_2, \quad \mathrm{E} = 2\mathrm{X}^T_0\mathrm{X}_0 + \mathrm{X}^T_1\mathrm{X}_2 + \mathrm{X}^T_2\mathrm{X}_1, \\
& h(\rho) = (\boldsymbol{\theta}_{LS}(\rho) - \boldsymbol{\theta}_0)^T(\kappa^{-1}\mathrm{V} + (1-\rho^2)(\mathrm{D} - \rho\mathrm{E})^{-1})^{-1}(\boldsymbol{\theta}_{LS}(\rho) - \boldsymbol{\theta}_0).
\end{aligned}$$

Hence, given a fixed value of ρ, we can estimate $\boldsymbol{\theta}$ by

$$\widehat{\boldsymbol{\theta}}(\rho) = (\kappa\mathrm{V}^{-1} + \mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1}(\kappa\mathrm{V}^{-1}\boldsymbol{\theta}_0 + \mathrm{X}^T\mathrm{U}^{-1}\mathrm{X}\boldsymbol{\theta}_{LS}(\rho)).$$

Note that $\widehat{\boldsymbol{\theta}}(\rho)$ does not depend on σ^2. Thus, even if σ^2 is unknown, we will obtain the same estimator. However, $\widehat{\boldsymbol{\theta}}(\rho)$ depends on ρ. Under the squared error loss, the estimator of $\boldsymbol{\theta}$ is given by $\widehat{\boldsymbol{\theta}} = E^{\pi_3(\rho|u)}[\widehat{\boldsymbol{\theta}}(\rho)]$. This expectation can be computed using Monte Carlo integration technique. Since u depend on z, which is random, we use EM algorithm (see Dempster, Laird and Rubin, 1977) to find $\widehat{\boldsymbol{\theta}}$ as

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + E^{\pi_3(\rho|\mathrm{u})}[(\kappa\mathrm{V}^{-1} + \mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1}\mathrm{X}^T\mathrm{U}^{-1}](\bar{\mathrm{u}} - \mathrm{X}\boldsymbol{\theta}_0),$$

where $\bar{\mathrm{u}} = \frac{1}{m}\sum_{i=1}^m \mathrm{u}(\mathrm{z}_i) = \mathrm{y} + \frac{1}{m}\sum_{i=1}^m \mathrm{z}_i = \mathrm{y} + \bar{\mathrm{z}}$.

In model selection (in terms of inclusion of the covariates), we are interested to test

$$H_0 : \boldsymbol{\theta}_{(2)} = 0 \quad \text{against} \quad H_1 : \boldsymbol{\theta}_{(2)} \neq 0,$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}^T_{(1)}, \boldsymbol{\theta}^T_{(2)})^T$, to decide whether we would include the covariates corresponding to $\boldsymbol{\theta}_{(2)}$ in the model or not. Hence, if the above H_0 is true, then

$$\boldsymbol{\theta}_{LS}(\rho) = \begin{pmatrix} \boldsymbol{\theta}_{LS(1)}(\rho) \\ \boldsymbol{\theta}_{LS(2)}(\rho) \end{pmatrix} \sim N_p\left(\begin{pmatrix} \boldsymbol{\theta}_{(1)} \\ \boldsymbol{\theta}_{(2)} = 0 \end{pmatrix}, \sigma^2(\mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1}\right).$$

Writing $(\mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1} = \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^T & \mathrm{C} \end{pmatrix}$, we have

$$\boldsymbol{\theta}_{LS(1)}(\rho)|\boldsymbol{\theta}_{LS(2)}(\rho), \rho, \sigma^2 \sim N_{p_1}(\boldsymbol{\theta}_{(1)} + \mathrm{B}\mathrm{C}^{-1}\boldsymbol{\theta}_{LS(2)}(\rho), \sigma^2\mathrm{F}),$$

where $\mathrm{F} = \mathrm{A} - \mathrm{B}\mathrm{C}^{-1}\mathrm{B}^T$. One need to take expectation with respect to σ^2 and ρ, and that can be tackled by Monte Carlo integration technique. Using standard technique, if $\boldsymbol{\theta}_{(1)} \sim N_{p_1}\left(\boldsymbol{\theta}_{0(1)}, \frac{\sigma^2}{\kappa}\mathrm{V}_{(1)}\right)$, the distribution of $\boldsymbol{\theta}_{LS(1)}$ given σ^2 and ρ is

$$\boldsymbol{\theta}_{LS(1)}(\rho)|\sigma^2, \rho \sim N_{p_1}(\boldsymbol{\theta}_{0(1)} + \mathrm{B}\mathrm{C}^{-1}\boldsymbol{\theta}_{LS(2)}(\rho), \sigma^2\mathrm{G}),$$

where $\mathrm{G} = \mathrm{F}^{-1} - \mathrm{F}^{-1}(\kappa \mathrm{V}_{(1)}^{-1} + \mathrm{F}^{-1})^{-1}\mathrm{F}^{-1}$. Let

$$m_0(\mathrm{u}) = E^{\pi_3(\rho|\mathrm{u})}[m_{0,1}(\boldsymbol{\theta}_{LS(1)}(\rho)|\boldsymbol{\theta}_{LS(2)}(\rho)) \cdot m_{0,2}(\boldsymbol{\theta}_{LS(2)}(\rho))]$$

be the marginal density of u under the null hypothesis. Then

$$\begin{aligned} m_0(\mathrm{u}) = E^{\pi_3(\rho|\mathrm{u})}\Bigg[& \frac{1}{(2\pi\sigma^2)^{p/2}|\mathrm{G}|^{1/2}|\mathrm{C}|^{1/2}} \\ & \times \exp\Bigg\{ -\frac{1}{2\sigma^2}[\boldsymbol{\theta}_{LS(2)(\rho)}^T \mathrm{C}^{-1}\boldsymbol{\theta}_{LS(2)(\rho)} \\ & + (\boldsymbol{\theta}_{LS(1)}(\rho) - \boldsymbol{\theta}_{0(1)} - \mathrm{BC}^{-1}\boldsymbol{\theta}_{LS(2)}(\rho))^T \mathrm{G}^{-1} \\ & \times (\boldsymbol{\theta}_{LS(1)}(\rho) - \boldsymbol{\theta}_{0(1)} - \mathrm{BC}^{-1}\boldsymbol{\theta}_{LS(2)}(\rho))] \Bigg\} \Bigg]. \end{aligned} \tag{2}$$

Write the marginal of u under the alternative hypothesis as

$$\begin{aligned} m_1(\mathrm{u}) = E^{\pi_3(\rho|\mathrm{u})}\Bigg[& \frac{|\mathrm{X}^T\mathrm{U}^{-1}\mathrm{X}|^{1/2}}{(2\pi\sigma^2)^{r/2}} \\ & \times \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{\theta}_{LS}(\rho) - \boldsymbol{\theta}_0)^T \mathrm{X}^T\mathrm{U}^{-1}\mathrm{X}(\boldsymbol{\theta}_{LS}(\rho) - \boldsymbol{\theta}_0) \right\} \Bigg]. \end{aligned}$$

Then we accept H_0 if $m_0(\mathrm{u})/m_1(\mathrm{u}) > 1$ and accept H_1 otherwise.

In Angers and Biswas (2001), several models are tried starting from one component equal to zero to all but one equal to zero. To implement, the standardized Bayesian estimates are ordered in the following way. Writing $Q_{(i)}$ as the ith diagonal element of the square matrix Q, we write

$$\mu_i = \frac{|\widehat{\theta}_i|}{E^{\pi_3(\rho|\mathrm{u})}[\sigma^2(\kappa\mathrm{V}^{-1} + \mathrm{X}^T\mathrm{U}^{-1}\mathrm{X})^{-1}]_{(i)}},$$

and suppose $\mu_{(1)} \leq \mu_{(2)} \leq \cdots \leq \mu_{(p)}$ be the ordered arrangement. The different possible models are then $M_l : \mu_{(1)} = \cdots = \mu_{(l)} = 0$ for $l = 1, 2, \cdots, p-1$. If m_l denotes the marginal probability of M_l, and if $B_l = m_l/m_0$, with m_0 being the marginal probability of the full model, then we accept M_{l^*} as the correct model for our purpose if

$$B_{l^*} = \max_{0 \leq l \leq p-1} B_l.$$

Note that here $B_0 = 1$.

It is to be noted that with the chosen *a priori* model, only numerical integration with respect to ρ is needed. In order to evaluate $\widehat{\boldsymbol{\theta}}$, the Monte Carlo method with importance sampling is used. The importance sampling function used to generate the ρ values is

$$g(\rho) \propto (1 - \rho^2)^{(\alpha+n)/2},$$

and 5000 iterations were made. For each value of ρ, the $\bar{\mathrm{z}}$ is computed using 1000 iterations. In practice, it is generated from $\bar{\mathrm{z}} \sim N_2(0, \sigma_Z^2 \mathrm{I}_\rho/1000)$.

>From the computation (*cf.* Table 4 and 5), we observe that except the constant term RSRBL (the effect of the retinopathy scale of the right eye of the baseline data as covariate for the left eye of the current study), RSLBL (the effect of the retinopathy scale of the left eye of the baseline data as covariate for the left eye of the current study) and GH came out as important covariates. In fact, Klein *et at.* (1988) also observed GH as an important covariate for retinopathy. Again the same scenario was observed in the analyses of Kim (1995) and Biswas and Das (2001). The analysis was carried out with these covariates only, for different κ.

3 Analysis of Concatenated Univariate Data

One can easily understand that an analysis with bivariate ordinal categorical data set is quite complicated and computationally expensive too. Moreover, one has to take care of the polychoric correlation (correlation between two categorical random variables). In the present study our objective is to see whether we can suitably transform the bivariate ordinal categorical data to an univariate ordinal categorical data in the sense that we can clearly catch the major features of the data in such transformed univariate data, whose analysis is quite simpler.

Table 1: Correlation coefficient between RS and Y_L, Y_R and Z

Variables	Correlation
RS and Y_L	0.9517
RS and Y_R	0.9506
RS and Z	0.9833

The first thing to do is, of course, to make a suitable transformation maintaining the ordinal nature and flavor. In order to do so, let

$$Z_i = aY_{Li} + bY_{Ri}.$$

We want to choose a and b such that the Z_i's are as similar as possible to the Y_{Li}'s and Y_{Ri}'s. This can be achieved by maximizing the sum of the correlation coefficient between Z and Y_L and that of Z and Y_R. It can be shown that

$$\rho_{Z,Y_L} + \rho_{Z,Y_R} = \frac{(1+\rho_{L,R})(a\sqrt{V_L} + b\sqrt{V_R})}{\sqrt{a^2V_L + b^2V_R + 2ab\rho_{L,R}\sqrt{V_LV_R}}}, \tag{3}$$

where

$$\begin{aligned} V_L &= \mathbb{V}\text{ar}(Y_{Li}), \\ V_R &= \mathbb{V}\text{ar}(Y_{Ri}), \\ \rho_{L,R} &= \mathbb{C}\text{orr}(Y_{Li}, Y_{Ri}). \end{aligned}$$

Equation (3) is maximum when

$$a = b\sqrt{\frac{V_R}{V_L}}.$$

If we impose the constraint that $\mathbb{V}\text{ar}(Z_i) = (\mathbb{V}\text{ar}(Y_{Li}) + \mathbb{V}\text{ar}(Y_{Li}))/2$, then we obtain

$$\begin{aligned} a &= \frac{1}{2}\sqrt{\frac{V_L + V_R}{V_L(1+\rho_{L,R})}}, \\ b &= \frac{1}{2}\sqrt{\frac{V_L + V_R}{V_R(1+\rho_{L,R})}}, \end{aligned}$$

and $\rho_{Z,Y_L} + \rho_{Z,Y_R} = \sqrt{2(1+\rho_{L,R})}$. Consequently, the "best" univariate transformation is given by

$$Z_i = \frac{1}{2}\sqrt{\frac{V_L + V_R}{1+\rho_{L,R}}}\left[\frac{\sqrt{V_R}Y_{Li} + \sqrt{V_L}Y_{Ri}}{\sqrt{V_LV_R}}\right]. \tag{4}$$

For the data set under consideration, the different quantities of equation (4) have been estimated from the data and the coefficient of Y_{Li} is $a = 0.5163$, and the one for Y_{Ri} is $b = 0.5174$. Consequently, Z_i is almost equal to the average between Y_{L_i} and Y_{Ri}. The correlation coefficients of Z with Y_L and Y_R are respectively $\rho_{Z,Y_L} = 0.9675$ and $\rho_{Z,Y_R} = 0.9672$.

An alternative univariate transformation of the bivariate data is done by the experimenters themselves. This is a retinopathy scale (RS) in which the retinopathy levels of both eyes are concatenated into a person-level scale. The RS used in the present work is a more current one than the one used in some earlier works by different authors. In finding RS, the worse eye is given greater weight. The fellow eye has either the same level or a lower level. All levels of proliferative retinopathy (60–85) are grouped together. This results in a 15-level scale: 10/10, 21/ < 21, 21/21, 31/ < 31, 31/31, 37/ < 37, 37/37, 43/ < 43, 43/43, 47/ < 47, 47/47, 53/ < 53, 53/53, 60 + / < 60+, 60 + /60+, which are numbered 0 through 14. The ordinal nature is maintained by simply putting arbitrarily larger weight to the worse eye (see Klein, Davis, Segal, Long, Harris, Haug, Magli and Syrjala, 1984). The correlation coefficients between the retinopathy scale and the other variables (Y_L, Y_R and Z) are given in Table 1. Hence Z is very similar (up to a linear transformation) to RS since is has a strong correlation with RS. However, the correlation of Z with Y_L and Y_R is slightly higher than the one of RS with Y_L and Y_R. The descriptive statistics of Z (along with the ones of Y_L and Y_R) are given in Table 2. >From this table, it can be seen that Z is very similar to Y_L and Y_R.

Let z be the vector of responses from all the individuals. We model z as

$$\text{z} = \tilde{\text{X}}^*\boldsymbol{\gamma} + \boldsymbol{\epsilon}^*,$$

Table 2: Descriptive statistics of Z, Y_L and Y_r

	Z	Y_L	Y_R
mean	30.571	29.549	29.599
std dev.	16.193	16.209	16.177
minimum	10.337	10	10
1^{st} quartile	16.882	21	21
median	26.882	31	31
3^{rd} quartile	38.254	37	37
maximum	73.395	71	75

where $\tilde{\mathrm{X}}^* = (\mathrm{X}_0\ \mathrm{X}_1\ \mathrm{X}_2)$ and $\boldsymbol{\gamma} = (\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$. A normal prior, as in (1), for $\boldsymbol{\gamma}$ is considered as

$$\boldsymbol{\gamma} \sim N_p(\boldsymbol{\gamma}_0, \frac{\sigma^2}{\kappa}\mathrm{V}),$$

with σ^2, κ and V may be different from the bivariate case. But in the univariate set up, no correlation parameter comes into consideration. Thus $\boldsymbol{\gamma}_0$ is estimated using z^*, $\tilde{\boldsymbol{X}}^*$ (baseline data) as

$$\hat{\boldsymbol{\gamma}}_0 = (\mathrm{X}^{*T}\mathrm{X}^*)^{-1}\mathrm{X}^{*T}\mathrm{z}.$$

Using standard technique, after some routine steps, it can be shown that the conditional posteriors of $\boldsymbol{\gamma}$ and σ^2 are:

$$\begin{aligned}
\boldsymbol{\gamma} \mid \mathrm{z}, \sigma^2 \sim N_p\Bigg(& \left[\kappa\mathrm{V}^{-1} + \tilde{\mathrm{X}}^T\tilde{\mathrm{X}}\right]^{-1}\left[\kappa\mathrm{V}^{-1}\boldsymbol{\gamma}_0 + \tilde{\mathrm{X}}^T\tilde{\mathrm{X}}\boldsymbol{\gamma}_{LS}\right], \\
& \qquad \sigma^2\left[\kappa\mathrm{V}^{-1} + \tilde{\mathrm{X}}^T\tilde{\mathrm{X}}\right]^{-1}\Bigg), \\
\sigma^2 \mid \mathrm{z} \sim I\Gamma\Bigg(& \frac{\alpha+n-2}{2}, 0.5 \times \Big[\beta + \mathrm{r}^T\mathrm{r} \\
& + (\boldsymbol{\gamma}_{LS} - \boldsymbol{\gamma}_0)^T\left[\kappa^{-1}\mathrm{V} + \left(\tilde{\mathrm{X}}^T\tilde{\mathrm{X}}\right)^{-1}\right]^{-1}(\boldsymbol{\gamma}_{LS} - \boldsymbol{\gamma}_0)\Big]\Bigg),
\end{aligned}$$

where $\boldsymbol{\gamma}_{LS} = \left(\tilde{\mathrm{X}}^T\tilde{\mathrm{X}}\right)^{-1}\tilde{\mathrm{X}}^T\mathrm{z}$, and $\mathrm{r} = \mathrm{z} - \mathrm{X}\boldsymbol{\gamma}_{LS}$. Hence, under the squared-error loss, the Bayes estimator of $\boldsymbol{\theta}$ is independent of σ^2 and it is given by:

$$\widehat{\boldsymbol{\gamma}} = \left[\kappa\mathrm{V}^{-1} + \tilde{\mathrm{X}}^T\tilde{\mathrm{X}}\right]^{-1}\left[\kappa\mathrm{V}^{-1}\gamma_0 + \tilde{\mathrm{X}}^T\tilde{\mathrm{X}}\boldsymbol{\gamma}_{LS}\right]$$

and its posterior expected loss is $\widehat{\sigma}^2\left[\kappa\mathrm{V}^{-1} + \tilde{\mathrm{X}}^T\tilde{\mathrm{X}}\right]^{-1}$, where

$$\begin{aligned}
\widehat{\sigma}^2 = \frac{1}{\alpha+n-4} \times \Big[& \beta + \mathrm{r}^T\mathrm{r} \\
& + (\boldsymbol{\gamma}_{LS} - \boldsymbol{\gamma}_0)^T\left[\kappa^{-1}\mathrm{V} + (\tilde{\mathrm{X}}^T\tilde{\mathrm{X}})^{-1}\right]^{-1}(\boldsymbol{\gamma}_{LS} - \boldsymbol{\gamma}_0)\Big].
\end{aligned}$$

The model selection process is similar to Section 2. However, since there is no correlation coefficient ρ to account for, the formulae are somehow easier. The marginal density of z under the hypothesis $H_0 : \boldsymbol{\theta}_{(2)} = 0$ is given by:

$$\begin{aligned}
m_0(\mathrm{z}) = & \frac{\Gamma([\alpha + p_{(1)}]/2]}{\Gamma(\alpha/2)\pi^{p_{(1)}/2}|\mathrm{A}_{(1)}|^{1/2}} \\
& \times \frac{\beta^{\alpha/2}}{\left[\beta + \mathrm{r}_{(1)}^T\mathrm{r}_{(1)} + \left(\boldsymbol{\gamma}_{LS(1)} - \boldsymbol{\gamma}_{0(1)}\right)^T\mathrm{A}_1^{-1}\left(\boldsymbol{\gamma}_{LS(1)} - \boldsymbol{\gamma}_{0(1)}\right)\right]^{(\alpha+p_{(1)})/2}},
\end{aligned}$$

where

$$\begin{aligned}
p_{(1)} &= \text{ the number of components in } \boldsymbol{\gamma}_{(1)}, \\
\mathrm{A}_{(1)} &= \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{1,2}^T, \quad \mathrm{r}_{(1)} = \mathrm{z} - \mathrm{X}_{(1)}\boldsymbol{\gamma}_{LS(1)},
\end{aligned}$$

and $\Sigma_{1,1}$, $\Sigma_{1,2}$, $\Sigma_{2,2,}$ are such that

$$\kappa \mathrm{V}^{-1} + \left(\tilde{\mathrm{X}}^T \tilde{\mathrm{X}}\right)^{-1} = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{1,2}^T & \Sigma_{2,2} \end{pmatrix}.$$

The marginal density under the full model is

$$m_1(\mathrm{z}) = \frac{\Gamma([\alpha+p]/2]}{\Gamma(\alpha/2)\pi^{p/2}|\kappa \mathrm{V}^{-1} + \left(\tilde{\mathrm{X}}^T \tilde{\mathrm{X}}\right)^{-1}|^{1/2}} \times \frac{\beta^{\alpha/2}}{\left[\beta + \mathrm{r}^T \mathrm{r} + (\boldsymbol{\gamma}_{LS} - \boldsymbol{\gamma}_0)^T \left[\kappa \mathrm{V}^{-1} + (\tilde{\mathrm{X}}^T \tilde{\mathrm{X}})^{-1}\right]^{-1} (\boldsymbol{\gamma}_{LS} - \boldsymbol{\gamma}_0)\right]^{(\alpha+p)/2}}.$$

To choose the best model, we follow the technique discussed at the end of Section 2.

4 Numerical Calculations and Conclusions

The 4-year follow-up WESDR data, on 629 individuals, are analyzed. The details of the computations in the bivariate case, can be obtained in Angers and Biswas (2001).

Table 3 presents the estimated values for the parameters of the full model, with all the covariates, for several values of κ, both for the bivariate and the univariate data as the two entries in each cell. Quite naturally, the parameters corresponding to RSRBR and RSLBR are not in the univariate model (represented by NA in Table 3). These covariates were used in both X_1 and X_2 in order to measure their influence separately on the retinophaty scale of each eye.

>From Table 3, it can be seen that the coefficient of RSRBR (RSRBL) is similar to the one of RSLBL (RSLBR) while using the bivariate ordinal data. Hence it can be concluded that the retinophaty scale of the right eye in baseline study has a similar effect of the retinopathy scale of the right eye in the current study. It can also be seen from this table that the estimated values for the covariates for both approach are not influenced by the choice of κ, mainly because the data set is quite large. For the full model, the correlation coefficient between the observations on both eyes and the predicted values is 0.786 for all values of κ in case of the bivariate data. Using the univariate transformation, the correlation coefficient between the observations and the predicted values increases to 0.819 for all values of κ. Furthermore, it can be seen that the regression coefficient of the variables which are person specific do not change much under the uni- and bivariate models. The variables RSRBR and RSLRL have a similar behavior. However, the three eyes specific variables, except left ME, change dramatically from one model to the other.

In Table 4, the estimated values of the coefficients of the covariates, along with the correlation between the observations and the predicted values for the "best" model are given. The "best" model is defined as the model maximizing the marginal probability of the observations. The "best" model for the univariate transformation involves only the baseline retinopathy scale (variable RSLBL). The estimated values does not depend on κ and it is equal to 1.130. In this set up, the correlation coefficient is 0.736. The second best model, with a correlation of 0.745, involves the baseline retinopathy scale (0.777) and the GH (1.208) for all values of κ.

Since it is quite difficult to compare all possible models, the models tested are obtained in the following way: (1) Compute the posterior mean and variance for all covariate in the model; (2) Compute the ratio of the posterior mean over the posterior standard deviation; (3) Delete the covariate with the smallest ratio from the model and repeat steps 1 to 3.

>From Table 4, it can be seen that the choice of the best bivariate model depends heavily on the choice of κ while it is always the same for the univariate set up. However, the covariates RSRBL, RSLBL and GH are included in almost all model, and hence we decided to fit our model with the covariates RSRBL, RSLBL, GH and a constant term. This model is adjusted in Table 5 for several values of κ. >From this table, it can be seen that the constant term is a decreasing function of κ, while the coefficient of GH is an increasing one. The estimated values of the coefficients of RSRBL and RSLBL do not depend on the choice of κ. The correlation between the observations and the predicted values is constant as a function of κ and is equal to 0.721.

The present paper is an attempt to analyze bivariate ordinal data using a general linear model in a Bayesian framework. Two very general and flexible models are proposed, one keeping the bivariate ordinal set up while the second one using a transformed univariate set up. The prior can be made as noninformative (or informative) by choosing suitable values for the hyperparameters κ, α and γ. The observations from the baseline study was used to elicit the prior mean θ_0 of the covariates.

Table 3: Results for the full model.

Parameter	$\kappa = 1$	$\kappa = 0.75$	$\kappa = 0.5$	$\kappa = 0.25$	$\kappa = 0.1$
Constant	-19.280	-19.664	-20.100	-20.450	-20.700
	-18.037	-18.695	-19.417	-20.251	-20.827
Right ME	11.777	11.786	11.800	11.791	11.800
	4.881	4.776	4.671	4.555	4.483
Right RE	-0.172	-0.173	-0.172	-0.174	-0.174
	0.192	0.194	0.197	0.200	0.203
Right IOP	-0.006	-0.003	0.001	0.003	0.004
	-0.290	-0.289	-0.288	-0.287	-0.286
RSRBR	0.366	0.366	0.366	0.366	0.366
	0.324	0.325	0.325	0.325	0.352
RSLBR	0.278	0.278	0.278	0.277	0.277
	NA	NA	NA	NA	NA
Left ME	13.399	13.399	13.300	13.403	13.500
	12.122	12.155	12.188	12.223	12.245
Left RE	0.126	0.125	0.127	0.124	0.123
	-0.246	-0.249	-0.253	-0.258	-0.260
Left IOP	-0.013	-0.011	-0.007	-0.004	-0.003
	0.240	0.244	0.249	0.254	0.257
RSRBL	0.293	0.292	0.293	0.292	0.292
	NA	NA	NA	NA	NA
RSLBL	0.361	0.361	0.360	0.360	0.360
	0.322	0.321	0.321	0.321	0.321
DuD	0.156	0.157	0.158	0.159	0.159
	0.161	0.162	0.163	0.164	0.165
GH	0.942	0.947	0.956	0.962	0.966
	0.928	0.939	0.952	0.966	0.977
SBP	0.006	0.006	0.006	0.007	0.007
	0.009	0.010	0.011	0.013	0.014
DBP	0.123	0.124	0.125	0.125	0.126
	0.119	0.120	0.121	0.123	0.124
BMI	0.340	0.342	0.344	0.347	0.349
	0.356	0.360	0.365	0.370	0.373
PR	0.061	0.062	0.063	0.064	0.065
	0.059	0.061	0.063	0.065	0.066
UP	-0.268	-0.290	-0.306	-0.327	-0.345
	-0.397	-0.425	-0.452	-0.484	-0.503
DI	-0.004	0.017	0.041	0.063	0.078
	-0.140	-0.099	-0.055	-0.004	0.030

Table 4: Results for the best model chosen using the maximum marginal probability and the bivariate ordinal observations.

Parameter	$\kappa = 1$	$\kappa = 0.75$	$\kappa = 0.5$	$\kappa = 0.25$	$\kappa = 0.1$
Constant	—	—	—	—	2.165
RSRBR	0.732	0.744	—	—	—
RSLBR	0.377	0.745	—	—	—
RSRBL	—	—	1.090	1.090	0.741
RSLBL	0.372	—	1.100	1.100	0.746
GH	—	1.173	—	—	0.977
DBP	0.156	—	—	—	—
Correlation	0.734	0.721	0.712	0.712	0.721

Table 5: Results for the chosen bivariate model.

Parameter	$\kappa = 1$	$\kappa = 0.75$	$\kappa = 0.5$	$\kappa = 0.25$	$\kappa = 0.1$
Constant	1.873	1.963	2.023	2.115	2.166
RSRBL	0.742	0.742	0.742	0.742	0.741
RSLBL	0.747	0.747	0.747	0.747	0.746
GH	1.003	0.994	0.989	0.981	0.977

Noninformative prior for σ^2 is used ($\alpha = 1$ and $\gamma = 0$), and a sensitivity analysis for the choice of κ is conducted. It can be seen that κ does not have a significant influence the resulting estimates. The model selection approach provides an opportunity to deal with the significant covariates only. It is observed that although a lot of covariates were recorded, only a few of them have significant contribution to the severity of retinopathy. Note that any other suitable standard criterion like the Bayes information criteria (BIC) could be used for model selection. It can also be seen that the results obtained using the univariate transformation are slightly better than those using the bivariate ordinal data because it has a slightly better correlation with the observation.

In the perspective of the WESDR study it could be of interest to analyze the 10 year follow-up data also, by a similar technique. But we could not access the data in the form of raw data.

Acknowledgments

The work is partially supported by a grant from NSERC. A part of the work of the second author was carried out when he was visiting the Département de mathématiques et de statistique, Université de Montréal. The second author thanks the department for its hospitality. The authors wish to thank Drs. Ronald Klein and Barbara E. K. Klein of the University of Wisconsin for providing both the baseline and 4-year follow-up data of the WESDR study. The WESDR project was originally supported in part by grant EY 03083 (R. Klein) from the National Eye Institute, NIH.

References

[1] Angers, J.-F. and Biswas, A., 2001: A Bayesian analysis of the four-year follow-up data of the Wisconsin epidemiologic study of diabetic retinopathy. Technical Report CRM-2757, Université de Montréal.

[2] Biswas, A. and Das, K., 2002: A Bayesian analysis of bivariate ordinal data: Wisconsin epidemiologic study of diabetic retinopathy revisited. *Statistics in Medicine*, **21**, 549–559.

[3] Dale, J.R., 1986: Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917.

[4] Das, K. and Biswas, A., 2001: Dirichlet process mixed model for bivariate ordered categorical data with application to the Wisconsin epidemiologic study of diabetic retinopathy. Technical Report. Applied Statistics Division, Indian Statistical Institute.

[5] Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977: 'Maximum likelihood estimation from incomplete data via the EM algorithm.' *Journal of the Royal Statistical Society, series B*, **39**, 1–22.

[6] Kim, K., 1995: A bivariate cumulative probit regression model for ordered categorical data. *Statistics in Medicine* , **14**, 1341–1352.

[7] Kim K., Lipsitz S.R. and Williamson, J.M., 1996: Regression models for bivariate ordered categorical data from ophthalmological studies. In Koo, J. Y., Park, B. U., Lee, K. W., Jeon, J. W. (eds). Collected Papers in Honor of Retirement of Professor Chung Han Yung. Seoul National University, Department of Computer Science and Statistics Alumni Association, pp. 36–55.

[8] Klein, B.E.K., Davis, M.D., Segal, P., Long, J.A., Harris, W.A., Haug, G.A., Magli, Y. and Syrjala, S., 1984: Diabetic retinopathy: assessment of severity and progression. *Ophthalmology* **91**, 10–17.

[9] Klein, R., Klein, B.E.K., Moss, S.E. and Cruickshanks, K.J., 1994: The Wisconsin epidemiologic study of diabetic retinopathy XIV. Ten-year incidence and progression of diabetic retinopathy. *Archives of Ophthalmology*, **112**, 1217–1228.

[10] Klein, R., Klein, B.E.K., Moss, S.E., Davis, M.D., DeMets, D.L., 1984: The Wisconsin Epidemiologic study of diabetic retinopathy, II: prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. *Archives of Ophthalmology*, **102**, 520–526.

[11] Klein, R., Klein, B.E.K., Moss, S.E., Davis, M.D. and DeMets, D.L., 1984: The Wisconsin Epidemiologic study of diabetic retinopathy, II: prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. *Archives of Ophthalmology*, **102**, 527–532.

[12] Klein, R., Klein, B.E.K., Moss, S.E., Davis, M.D. and DeMets, D.L., 1988: Glycosylated hemoglobin predicts the incidence and progression of diabetic retinopsthy. *Journal of the American Medical Association 260*, 2864–2871.

[13] Klein, R., Klein, B.E.K., Moss, S.E., Davis, M.D. and DeMets, D.L., 1989: The Wisconsin epidemiologic study of diabetic retinopathy IX. Four-year incidence and progression of diabetic retinopathy when age at diagnosis is less than 30 years. *Archives of Ophthalmology*, **107**, 237–243.

[14] Klein, R., Klein, B.E.K., Moss, S.E., DeMets, D.L., Kaufman, I. and Voss, P.S., 1984: Prevalence of diabetes mellitus in southern Wisconsin. *American Journal of Epidemiology*, **119**, 54–61.

[15] Molenberghs, G. and Lesaffre, E., 1994: Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, **89**, 633–644.

[16] Williamson, J. and Kim, K., 1996: A global odds ratio regression model for bivariate ordered categorical data from ophthalmologic studies. *Statistics in Medicine*, **15**, 1507–1518.

[17] Williamson, J. M., Kim, K. and Lipsitz, S. R., 1995: Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, **90**, 1432–1437.

[18] Williamson, J., Lipsitz, S.R. and Kim, K., 1999: GEECAT and GEEGOR: computer programs for the analysis of correlated categorical response data. *Computer Methods and Programs in Biomedicine*, **58**, 25–34.