

Article

# Multivariate Mixed Response Model with Pairwise Composite-Likelihood Method

Hao Bai <sup>1,†</sup>, Yuan Zhong <sup>1,†</sup> , Xin Gao <sup>1,\*</sup>  and Wei Xu <sup>2</sup>

<sup>1</sup> Department of Mathematics and Statistics, York University, Toronto, ON M3J 1P3, Canada; baihaohao@gmail.com (H.B.); aqua.zhong@gmail.com (Y.Z.)

<sup>2</sup> Department of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5S 1A1, Canada; wei.xu@uhnres.utoronto.ca

\* Correspondence: xingao@mathstat.yorku.ca

† These authors contributed equally to this work.

Received: 30 June 2020; Accepted: 13 July 2020; Published: 15 July 2020



**Abstract:** In clinical research, study outcomes usually consist of various patients' information corresponding to the treatment. To have a better understanding of the effects of different treatments, one often needs to analyze multiple clinical outcomes simultaneously, while the data are usually mixed with both continuous and discrete variables. We propose the multivariate mixed response model to implement statistical inference based on the conditional grouped continuous model through a pairwise composite-likelihood approach. It can simplify the multivariate model by dealing with three types of bivariate models and incorporating the asymptotical properties of the composite likelihood via the Godambe information. We demonstrate the validity and the statistic power of the multivariate mixed response model through simulation studies and clinical applications. This composite-likelihood method is advantageous for statistical inference on correlated multivariate mixed outcomes.

**Keywords:** composite likelihood; multivariate analysis; mixed outcome; Godambe information

## 1. Introduction

Clinical research, such as toxicity studies and laboratory examinations, can provide relevant information for measuring the effect of various treatments or experiments on the patients. This type of research needs to jointly analyze different experimental outcomes, while the research outcomes collected during the treatment are correlated and mixed with both categorical and continuous variables. For example, we are studying the efficacy of treatments along with the toxicity and adverse drug reactions simultaneously. In this case, the severity level could be measured as discrete or ordinal data, while the clinical examination results such as the blood test measures are continuous. In the traditional approach, these multiple outcomes are analyzed by different linear models to estimate the effects of the treatments together with the relevant clinical and demographic information. However, this approach ignores the correlation between the outcomes and only provides marginal inferences. Thus, it is desirable to develop a multivariate approach, which can jointly model the multiple mixed-type responses with the treatment and clinical covariates.

In the recent literature, there are two main approaches to build a joint model for mixed response variables. The conditional Gaussian distribution model (CGDM) can decompose the joint distribution of mixed response variables into a combination of the conditional distribution and the marginal distribution. In the bivariate mixed case, the conditional Gaussian distribution model can produce a conditional distribution of the categorical variables given the continuous variables and marginal distribution for the continuous variables. In particular, Cox [1] provided

the logistic conditional distribution for binary variables, and Cox and Wermuth [2] extended the model with a probit-type function and showed the potential connection to the latent variable model. Another conditional Gaussian distribution model, referred to as the general location model (GLOM), was proposed by Olkin and Tate [3]. They adopted the opposite factorization, which consists of a conditional normal distribution given the categorical variables and marginal multinomial distribution. Teixeira-Pinto and Normand [4] compared this approach with the models proposed by Sammel et al. [5,6] in a comprehensive review. Yang et al. [7] extended the model to mixed Poisson and continuous response variables through a likelihood-based approach.

The grouped continuous model (GCM) proposed by Anderson and Pemberton [8] and de Leon [9] provides another solution for this problem. The fundamental technique allows the categorical variables to be treated as partitioned continuous latent variables with different nonoverlapping intervals (Poon and Lee [10]; Skrondal and Rabe-Hesketh [11]). This type of transformation allows the latent variables to follow a multivariate Gaussian distribution. Poon and Lee [10] demonstrated maximum likelihood estimation for latent variables with polychoric correlations. As an extension of the grouped continuous model, de Leon [9] proposed the conditional grouped continuous model (CGCM) to build a joint model for the variables mixed with categorical and continuous outcomes. Catalano and Ryan [12], Catalano [13], and Najita et al. [14] applied the conditional grouped continuous model to the studies of fetal toxicity for longitudinal data. Gueorguieva and Agresti [15] proposed that the estimation of correlated mixed response variables can be obtained by the expectation–maximization (EM) algorithm. Zhang et al. [16] extended this to the parameter expanded EM algorithm under the full-likelihood approach. de Leon and Carrière [17,18] developed a general mixed-data model, which aggregates the CGDM and CGCM to jointly analyze correlated nominal, ordinal, and continuous data together.

It remains computationally challenging to estimate the joint distribution of multivariate mixed data. The composite-likelihood method is based on compounded lower-dimensional distributions, which offers an alternative solution to the estimation problem (Lindsay [19], Cox and Reid [20], Varin [21], Varin et al. [22], and Reid et al. [23]). Faes et al. [24] applied this method for longitudinal data with mixed outcomes. In their model, the correlation structure is induced by the random effect, which does not have a closed-form expression. We aim to follow the approach of the CGCM and use the composite-likelihood method to analyze the joint distribution of multivariate mixed-type response variables, where the categorical response variables are modeled by continuous latent variables. The parameters of the mean structure, as well as the correlation among different outcomes, can be estimated simultaneously through the numerical algorithm. The proposed composite-likelihood method consists of three types of bivariate joint densities: two continuous outcomes; two discrete outcomes modeled by two continuous latent variables; and two mixed outcomes with one continuous and one categorical variable.

We provide the numerical algorithm for composite-likelihood estimation and discuss the asymptotical properties of the composite-likelihood estimates. In addition, we derive three composite-likelihood test statistics for joint inference on the multivariate mixed response model. Simulation studies were conducted to examine the empirical performance of the proposed method in comparison with the conventional approaches. The algorithm was applied to the clinical data from a colorectal cancer study. We analyze the effect of the treatment and other clinical factors' on multiple correlated responses of the patients.

## 2. Methodology

### 2.1. Model Setup

Suppose there are  $n$  observations  $z_1, z_2, \dots, z_i, \dots, z_n$  in a clinical dataset, and each observation contains  $q$  multiple outcomes  $z_i = (z_{i1}, z_{i2}, \dots, z_{ij}, \dots, z_{iq})^T$ , which are correlated and mixed with continuous and binary variables. Suppose we wish to model the effects of a collection of covariates, and the generalized linear model can be constructed for each outcome as

$$g_j(E(z_{ij})) = \mathbf{x}_{ij}^T \boldsymbol{\beta}_j,$$

in which the covariate  $\mathbf{x}_{ij}$  with  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, q$  for different responses can be the same or different, and  $g_j$  denotes the link function used for the  $j$ th response. If we want to analyze these multivariate mixed-type outcomes simultaneously with the corresponding covariates, the conventional likelihood-based approach needs to identify the multivariate joint densities of all responses. Alternatively, we can set up the pairwise likelihood function for responses  $z_{ij}$  and  $z_{ik}$  as

$$L_{jk}(\boldsymbol{\theta}_{jk}) = \prod_{i=1}^n f(z_{ij}, z_{ik}), \quad (1)$$

where  $\boldsymbol{\theta}_{jk}$  denotes the parameters associated with the pairwise likelihood. The log-likelihood function is given by  $l_{jk}(\boldsymbol{\theta}_{jk}) = \log L_{jk}(\boldsymbol{\theta}_{jk})$ , and the score function is given by

$$U_{jk}(\boldsymbol{\theta}_{jk}) = \sum_{i=1}^n f(z_{ij}, z_{ik})^{-1} \frac{\partial}{\partial \boldsymbol{\theta}_{jk}} f(z_{ij}, z_{ik}). \quad (2)$$

According to Lindsay [19] and Cox and Reid [20], the pairwise composite-likelihood function of these  $q$  response variables is the product of  $\binom{q}{2}$  paired likelihood functions:

$$\text{CL}(\boldsymbol{\theta}) = \prod_{j=1}^{q-1} \prod_{k=j+1}^q L_{jk}(\boldsymbol{\theta}_{jk}),$$

and the score function is constructed by differentiating the composite log-likelihood function:

$$U(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log \text{CL}(\boldsymbol{\theta}). \quad (3)$$

Our multivariate mixed response model estimates the parameters of interest  $\boldsymbol{\theta}$  by numerically solving the composite score Function (3) equal to 0 through the Newton–Raphson [25] method. Since the outcomes are mixed with continuous and categorical variables, the score Function (2) can be derived under three different bivariate structures: outcomes with both continuous response variables, outcomes with both binary response variables, and outcomes mixed with one continuous and one binary response variable.

### 2.1.1. Case 1: Continuous Outcomes

We first discuss the case that both responses  $z_{ij}$  and  $z_{ik}$  are continuous and assume they follow a bivariate normal distribution:

$$\begin{pmatrix} z_{ij} \\ z_{ik} \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \mu_{ij} \\ \mu_{ik} \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \rho_{jk} \sigma_j \sigma_k \\ \rho_{jk} \sigma_j \sigma_k & \sigma_k^2 \end{pmatrix} \right),$$

where the mean structure of  $z_{ij}$  and  $z_{ik}$  is associated with the generalized linear models as  $\mu_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_j$  and  $\mu_{ik} = \mathbf{x}_{ik}^T \boldsymbol{\beta}_k$ . Thus, the pairwise density of  $z_{ij}$  and  $z_{ik}$  is the same as a bivariate normal density based on Johnson and Wichern [26]:

$$f(z_{ij}, z_{ik}) = \frac{1}{2\pi\sigma_j\sigma_k\sqrt{1-\rho_{jk}^2}} \exp \left[ -\frac{1}{2(1-\rho_{jk}^2)} \times \left( \frac{(z_{ij} - \mu_{ij})^2}{\sigma_j^2} - \frac{2\rho(z_{ij} - \mu_{ij})(z_{ik} - \mu_{ik})}{\sigma_j\sigma_k} + \frac{(z_{ik} - \mu_{ik})^2}{\sigma_k^2} \right) \right].$$

The pairwise composite-likelihood function can be constructed by

$$L_{jk}(\theta_{jk}) = \prod_{i=1}^n f(z_{ij}, z_{ik}),$$

with the parameters of interest  $\theta_{jk} = \{\beta_j, \beta_k, \sigma_j, \sigma_k, \rho_{jk}\}$ , and the log-likelihood function  $l_{jk}(\theta_{jk}) = \sum_{i=1}^n \log f(z_{ij}, z_{ik})$ , which gives the score function of both continuous responses used in the Equation (2):

$$U_{jk}(\theta_{jk}) = \frac{\partial l_{jk}(\theta_{jk})}{\partial \theta_{jk}}. \tag{4}$$

### 2.1.2. Case 2: Binary Outcomes

For binary outcomes, the transformation can be obtained using the threshold value  $t$  such that if the latent variable  $z_{ij}^* \geq t$ , then the response  $z_{ij} = 1$ , otherwise  $z_{ij} = 0$ . Without loss of generality, the value of  $t$  is set to 0, and we have  $\mu_{ij} = P(z_{ij} = 1) = P(z_{ij}^* \geq 0)$ . We follow the grouped continuous model settings with the distributional assumptions  $z_{ij} \sim \text{Bernoulli}(\mu_{ij})$  and  $z_{ij}^* \sim N(\mu_{ij}^*, \sigma_j^2)$ . Thus, the model can be constructed with the covariates  $x_{ij}$  as

$$\text{probit}(\mu_{ij}) = x_{ij}^T \beta_j.$$

Since the mean  $\mu_{ij}^*$  and the variance  $\sigma_j^2$  are not identifiable at the same time, we follow the rescaling method based on Dunson [27] by setting  $\sigma_j^2 = 1$ . Therefore, the mean of the latent variable can be simplified as  $\mu_{ij}^* = x_{ij}^T \beta_j$ .

Since we jointly analyze binary variables  $z_{ij}$  and  $z_{ik}$ , the paired latent variables  $z_{ij}^*$  and  $z_{ik}^*$  are generated from a model with the polychoric correlation. Thus, the joint function of  $(z_{ij}, z_{ik})$  can be derived by following composite-likelihood functions in four cases

$$\begin{aligned} P(z_{ij} = 1, z_{ik} = 1) &= P(z_{ij}^* \geq 0, z_{ik}^* \geq 0) = \Phi_2(\mu_{ij}^*, \mu_{ik}^*, \rho_{jk}), \\ P(z_{ij} = 1, z_{ik} = 0) &= P(z_{ij}^* \geq 0, z_{ik}^* < 0) = \Phi_2(\mu_{ij}^*, -\mu_{ik}^*, -\rho_{jk}), \\ P(z_{ij} = 0, z_{ik} = 1) &= P(z_{ij}^* < 0, z_{ik}^* \geq 0) = \Phi_2(-\mu_{ij}^*, \mu_{ik}^*, -\rho_{jk}), \\ P(z_{ij} = 0, z_{ik} = 0) &= P(z_{ij}^* < 0, z_{ik}^* < 0) = \Phi_2(-\mu_{ij}^*, -\mu_{ik}^*, \rho_{jk}), \end{aligned}$$

where  $\pm\rho_{jk}$  represents the polychoric correlation under different scenarios and  $\Phi_2$  denotes the bivariate normal cumulative density function. The equations above can be rewritten as

$$\begin{aligned} P(z_{ij}, z_{ik}) &= \Phi_2((2z_{ij} - 1)\mu_{ij}^*, (2z_{ik} - 1)\mu_{ik}^*, (2z_{ij} - 1)(2z_{ik} - 1)\rho_{jk}) \\ &= \Phi_2(s_{ij}\mu_{ij}^*, s_{ik}\mu_{ik}^*, s_{ij}s_{ik}\rho_{jk}), \end{aligned}$$

with  $s_{ij} = (2z_{ij} - 1)$  and  $s_{ik} = (2z_{ik} - 1)$ . The log-likelihood function of paired binary responses is

$$l_{jk}(\theta_{jk}) = \sum_{i=1}^n \log \Phi_2(s_{ij}\mu_{ij}^*, s_{ik}\mu_{ik}^*, s_{ij}s_{ik}\rho_{jk}).$$

The score function can be derived as

$$U_{jk}(\theta_{jk}) = \sum_{i=1}^n P(z_{ij}, z_{ik})^{-1} \frac{\partial}{\partial \theta_{jk}} \log P(z_{ij}, z_{ik}). \tag{5}$$

### 2.1.3. Case 3: Mixed Outcomes

When one response variable  $z_{ij}$  is binary, and another response variable  $z_{ik}$  is continuous, we can model the mixed outcomes together by adopting a latent normal variable  $z_{ij}^*$  as the conditional grouped continuous model. Thus, they follow a bivariate normal distribution given the polyserial correlation  $\rho_{jk}$ :

$$\begin{pmatrix} z_{ij}^* \\ z_{ik} \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \mu_{ij}^* \\ \mu_{ik} \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \rho_{jk}\sigma_j\sigma_k \\ \rho_{jk}\sigma_j\sigma_k & \sigma_k^2 \end{pmatrix} \right).$$

The factorization method can be applied for the joint bivariate normal distribution with a marginal density for the continuous response  $P(z_{ik})$  and a conditional density for the latent variable  $P(z_{ij}^*|z_{ik})$  given the continuous response.

1.  $z_{ik}$  is the continuous response

$$z_{ik} \sim N(\mu_{ik}, \sigma_k^2);$$

2.  $z_{ij}^*$  is the latent variable for the binary response  $z_{ij}$ :

$$z_{ij}^*|z_{ik} \sim N\left(\mu_{ij|k} = \mu_{ij}^* + \rho_{jk} \frac{\sigma_j}{\sigma_k} (z_{ik} - \mu_{ik}), \sigma_{j|k}^2 = \sigma_j^2 (1 - \rho_{jk}^2)\right).$$

Let  $d_{ij} = -\frac{\mu_{ij|k}}{\sigma_{j|k}}$ , the conditional probability  $P(z_{ij}^*|z_{ik})$  can be written as

$$P(z_{ij}^*|z_{ik}) = \begin{cases} \Phi(d_{ij}), & \text{if } z_{ij} = 0, \\ \Phi(-d_{ij}), & \text{if } z_{ij} = 1. \end{cases}$$

The pairwise likelihood function and log-likelihood function is given as follows:

$$L_{jk} = \prod_{i=1}^n P(z_{ij}^*|z_{ik})p(z_{ik}),$$

$$l_{jk} = \sum_{i=1}^n \log P(z_{ik}) + \log \mathbf{1}(z_{ij} = 0)\Phi(d_{ij}) + \log \mathbf{1}(z_{ij} = 1)\Phi(-d_{ij}).$$

The score function can be obtained as

$$U_{jk} = \sum_{i=1}^n \frac{1}{P(z_{ik})} \frac{\partial P(z_{ik})}{\partial \theta_{jk}} + \left[ \frac{\mathbf{1}(z_{ij} = 0)}{\Phi(d_{ij})} - \frac{\mathbf{1}(z_{ij} = 1)}{1 - \Phi(d_{ij})} \right] \frac{\partial \Phi(d_{ij})}{\partial \theta_{jk}}. \tag{6}$$

A similar formulation was extended to the longitudinal settings by Najita et al. [14]. More derivation details are included in Appendix A.

The model setting above can simplify the score Function (3) with three types of components as in Equations (4)–(6). We can obtain the maximum composite-likelihood estimate  $\hat{\theta}$  (MCLE) for the parameters  $\theta$  by solving the score function numerically. Our numerical algorithm is presented below (Algorithm 1):

**Algorithm 1:** Multivariate Mixed Response Model Algorithm.

1. Select the initial values  $\theta^{(0)}$  of the parameters. Calculate the score function  $U(\theta^{(0)})$  in Equation (3) and the Hessian matrix  $\nabla_{\theta}U(\theta^{(0)})$  as the derivative of the score function;
2. Implement the Newton–Raphson method to obtain the updated parameters  $\theta^{(t)}$  with  $t = 0, 1, \dots$ ;

$$\theta^{(t+1)} = \theta^{(t)} - [\nabla_{\theta}U(\theta^{(t)})]^{-1}U(\theta^{(t)});$$

3. Calculate the score function and Hessian matrix with the updated parameters  $\theta^{(t+1)}$ ;
4. Repeat step 2 and 3 until convergence.

*2.2. Statistical Inference Using the Composite-Likelihood Method*

Cox and Hinkley [28] and Kent [29] presented various hypothesis testing procedures using the full-likelihood function. The composite-likelihood function can be treated as the misspecified-likelihood function. Its asymptotic properties were reviewed and discussed by Varin et al. [22], Reid et al. [23], Jin [30], and Gao and Song [31]. Following this framework, the pairwise composite-likelihood function implemented in our proposed model produces the estimators, which are consistent and asymptotically normally distributed.

The Godambe information [32]  $G$  of the parameters  $\theta$  for the log composite-likelihood function involves the sensitivity matrix  $H$  and the variability matrix  $J$ :

$$G(\theta) = H(\theta)J^{-1}(\theta)H(\theta),$$

where the sensitivity matrix and variability matrix are defined as

$$H(\theta) = E_{\theta}\{-\nabla_{\theta}U(\theta; z_i)\} \text{ and } J(\theta) = \text{Var}_{\theta}\{U(\theta; z_i)\},$$

where  $U(\theta; z_i)$  denotes the score function of the  $i$ th observation and the total score function  $U(\theta) = \sum_{i=1}^n U(\theta; z_i)$ .

**Theorem 1.** Let  $\theta_0 \in \mathbb{R}^d$  denote the true parameter value to set up the multivariate mixed response model. Under regularity conditions,  $n \rightarrow \infty$ , the maximum composite-likelihood estimator  $\hat{\theta}$  is asymptotically normally distributed as

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N_d(\mathbf{0}, G^{-1}).$$

**Theorem 2.** Under regularity conditions,  $n \rightarrow \infty$ , the maximum composite-likelihood estimator  $\hat{\theta}$  is consistent to  $\theta_0$  satisfying  $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$ .

The sensitivity matrix  $H$  and the variability matrix  $J$  can be evaluated by the empirical estimates under the maximum composite-likelihood estimators:

$$H(\hat{\theta}) = -\frac{1}{n} \sum_{i=1}^n \nabla_{\theta}U(\theta; z_i)|_{\hat{\theta}} \text{ and } J(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n U(\theta; z_i)U(\theta; z_i)^T.$$

Furthermore, according to Theorem 1, the composite Wald statistic, the composite score statistic, and the composite-likelihood ratio statistic for testing the null hypothesis  $H_0: \theta = \theta_0$  are given respectively by

$$\begin{aligned} W_e &= n(\hat{\theta} - \theta_0)G(\hat{\theta})(\hat{\theta} - \theta_0) \\ W_u &= n^{-1}U(\theta_0)J^{-1}(\hat{\theta})U(\theta_0), \\ W &= 2\{\log \text{CL}(\hat{\theta}) - \log \text{CL}(\theta_0)\}. \end{aligned}$$

### Testing with Nuisance Parameters

Suppose the parameters are partitioned as  $\theta = \{\psi, \lambda\}$  with  $\psi \in \mathfrak{R}^s$ ,  $\lambda \in \mathfrak{R}^p$ , and  $d = s + p$ . The parameter of interest is  $\psi$ , and  $\lambda$  is treated as the nuisance parameter for hypothesis testing. In this setting, the Godambe information matrix and its inverse can be partitioned as

$$G = \begin{bmatrix} G_{\psi\psi} & G_{\psi\lambda} \\ G_{\lambda\psi} & G_{\lambda\lambda} \end{bmatrix} \text{ and } G^{-1} = \begin{bmatrix} G^{\psi\psi} & G^{\psi\lambda} \\ G^{\lambda\psi} & G^{\lambda\lambda} \end{bmatrix},$$

and the inverse of the submatrix pertaining to  $\psi$  is given by  $G_{\psi\psi,\lambda} = (G^{\psi\psi})^{-1} = G_{\psi\psi} - G_{\psi\lambda}G_{\lambda\lambda}^{-1}G_{\lambda\psi}$ . According to the asymptotic theorem, the composite Wald statistics under the null hypothesis  $H_0: \psi = \psi_0$  using  $\hat{\lambda}(\psi_0)$  is given by

$$W_e(\psi_0) = n(\hat{\psi} - \psi_0)G_{\psi\psi,\lambda}(\hat{\psi} - \psi_0),$$

which has an asymptotic  $\chi_q^2$  distribution. Similarly, we define the composite score statistics:

$$W_u(\psi_0) = n^{-1}U(\psi_0, \hat{\lambda}(\psi_0))H^{\psi\psi}G_{\psi\psi,\lambda}H^{\psi\psi}U(\psi_0, \hat{\lambda}(\psi_0)),$$

where the matrix  $H^{\psi\psi}$  can be obtained by partitioning the inverse sensitivity matrix  $H$ . Furthermore, the composite-likelihood ratio statistic can be obtained by

$$W(\psi_0) = 2\{\log \text{CL}(\hat{\psi}, \hat{\lambda}) - \log \text{CL}(\psi_0, \hat{\lambda}(\psi_0))\},$$

with the unrestricted maximum composite-likelihood estimate  $\hat{\theta} = \{\hat{\psi}, \hat{\lambda}\}$ . However, the asymptotic distribution of the composite-likelihood ratio under  $H_0$  is given by  $\sum_{j=1}^q \lambda_j \chi_{1(j)}^2$ , where  $\chi_{1(j)}^2$  are independent  $\chi_1^2$  variates and  $\lambda_1, \lambda_2, \dots, \lambda_q$  are the eigenvalues of the matrix  $H_{\psi\psi,\lambda}G^{\psi\psi}$  with  $H_{\psi\psi,\lambda} = H_{\psi\psi} - H_{\psi\lambda}H_{\lambda\lambda}^{-1}H_{\lambda\psi}$ . There are different adjustments to this nonstandard weighted chi-square distribution (Rotnitzky and Jewell [33], Geys et al. [34], and Pace et al. [35]). For example, we can apply the adjustment by introducing the scaling factor  $\bar{\lambda} = \sum_{j=1}^q \lambda_j / q$ , then the adjusted composite-likelihood ratio has the same asymptotic distribution as  $W_e(\psi_0)$  and  $W_u(\psi_0)$ :

$$\frac{W}{\bar{\lambda}} \xrightarrow{d} \chi_s^2. \quad (7)$$

Therefore, the composite-likelihood method can simplify the modeling of correlated responses with multiple generalized linear models and allow users to conduct statistic inferences on parameters of interest from different generalized linear models. Moreover, we can select a subset of the parameters and conduct the further inferential assessment in the presence of nuisance parameters.

### 3. Simulation

Different simulation studies were implemented to show the validity of the multivariate mixed response model. The estimation results from the proposed model are compared with the full-likelihood and marginal approaches, respectively.

#### 3.1. Comparison with Maximum Full-Likelihood Estimation

In the multivariate regression with correlated continuous outcomes, the full-likelihood estimation can be conducted without numerical integration. Thus, we can compare the maximum composite-likelihood estimates with the full-likelihood approach through the simulation study. The simulated samples contain four continuous response variables  $z_{ic_1}, z_{ic_2}, z_{ic_3}$ , and  $z_{ic_4}$ , which are generated from Equation (8):

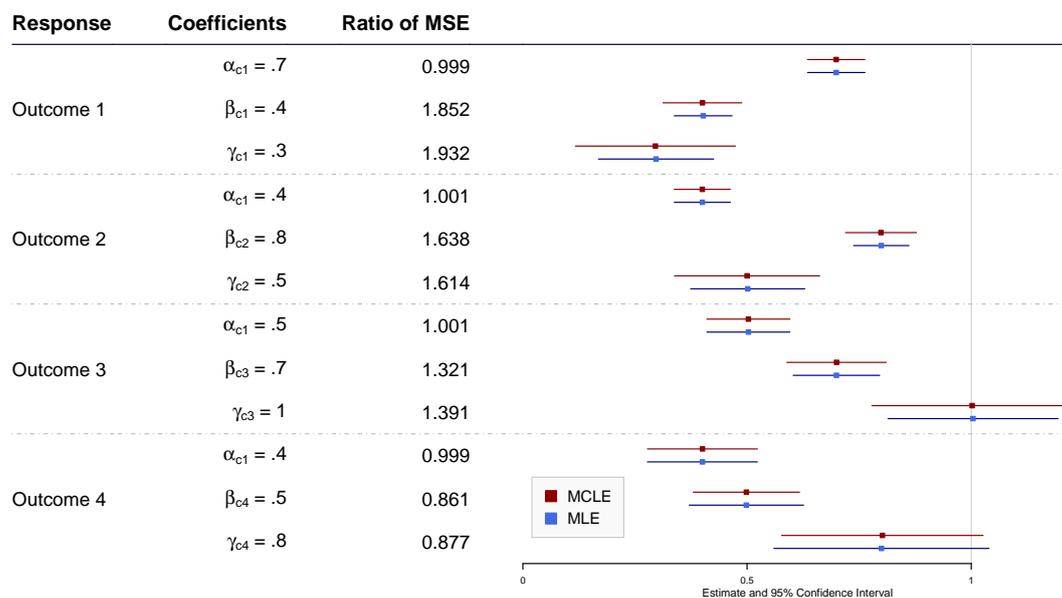
$$\begin{aligned}
 z_{ic_1} &= \alpha_{c_1} + \beta_{c_1}x_{ic_1} + \gamma_{c_1}y_{ic_1} + \varepsilon_{ic_1}, \\
 z_{ic_2} &= \alpha_{c_2} + \beta_{c_2}x_{ic_2} + \gamma_{c_2}y_{ic_2} + \varepsilon_{ic_2}, \\
 z_{ic_3} &= \alpha_{c_1} + \beta_{c_1}x_{ic_1} + \gamma_{c_1}y_{ic_1} + \varepsilon_{ic_1}, \\
 z_{ic_4} &= \alpha_{c_2} + \beta_{c_2}x_{ic_2} + \gamma_{c_2}y_{ic_2} + \varepsilon_{ic_2}.
 \end{aligned}
 \tag{8}$$

The covariates  $\{x_{ic_1}, x_{ic_2}, x_{ic_1}, x_{ic_2}\} \sim N(0, 1)$  and  $\{y_{ic_1}, y_{ic_2}, y_{ic_1}, y_{ic_2}\} \sim N(0, 0.5)$  are independently simulated. The errors are correlated and generated from a multivariate normal distribution  $N_4(\mathbf{0}, \Sigma)$ , and the variance–covariance matrix  $\Sigma$  is given by

$$\begin{bmatrix}
 \sigma_{c_1}^2 & \sigma_{c_1}\sigma_{c_2}\rho_{c_1c_2} & \sigma_{c_1}\sigma_{c_3}\rho_{c_1c_3} & \sigma_{c_1}\sigma_{c_4}\rho_{c_1c_4} \\
 \sigma_{c_1}\sigma_{c_2}\rho_{c_1c_2} & \sigma_{c_2}^2 & \sigma_{c_2}\sigma_{c_3}\rho_{c_2c_3} & \sigma_{c_2}\sigma_{c_4}\rho_{c_2c_4} \\
 \sigma_{c_1}\sigma_{c_3}\rho_{c_1c_3} & \sigma_{c_2}\sigma_{c_3}\rho_{c_2c_3} & \sigma_{c_3}^2 & \sigma_{c_3}\sigma_{c_4}\rho_{c_3c_4} \\
 \sigma_{c_1}\sigma_{c_4}\rho_{c_1c_4} & \sigma_{c_2}\sigma_{c_4}\rho_{c_2c_4} & \sigma_{c_3}\sigma_{c_4}\rho_{c_3c_4} & \sigma_{c_4}^2
 \end{bmatrix}.$$

In the simulation, the variances are designed as  $\sigma_{c_1}^2 = 1, \sigma_{c_2}^2 = 1, \sigma_{c_3}^2 = 2.25,$  and  $\sigma_{c_4}^2 = 4,$  and an identical correlation  $\rho = 0.3$  is applied between the errors in the data generating process.

The simulation results (Figure 1) were obtained through 1000 independent replications. The maximum composite-likelihood estimators demonstrate similar performance when compared with the full-likelihood approach. The simulated results also show that the estimates are close to each other, and the maximum likelihood estimators have slightly higher relative efficiency.



**Figure 1.** The comparison between the maximum full-likelihood estimation and maximum composite-likelihood estimation for the regression coefficients on the multivariate continuous outcomes. The ratio of the mean squared error (MSE) was computed using the MSE of the maximum composite-likelihood estimate (MCLE) over the MSE of the maximum likelihood estimate (MLE).

### 3.2. Comparison with the Marginal Approach

For the mixed outcome regression, the full-likelihood approach is computationally challenging, and marginal regression is often resorted to in order to conduct the analysis. We implemented simulation studies to evaluate the performance of our proposed method in comparison with marginal regression. We first tested the overall performance of the point estimates when the outcomes had

different levels of dependency and covariates. Next, we focused on the test of the composite statistics. The multivariate mixed response model can provide the statistical inference with nuisance parameters and attains a higher statistical power in terms of dealing with joint inference.

### 3.2.1. Simulation Settings

We generated the sample data consisting of two binary responses  $z_{ib_1}$  and  $z_{ib_2}$  and two continuous responses  $z_{ic_1}$  and  $z_{ic_2}$ . The binary variables were obtained based on the corresponding latent normal variables  $z_{ib_1}^*$  and  $z_{ib_2}^*$  through the probit link function:

$$\begin{aligned} \text{probit}(\mu_{ib_1}) &= \mu_{ib_1}^*, \\ \text{probit}(\mu_{ib_2}) &= \mu_{ib_2}^*. \end{aligned}$$

The simulation studies of the responses are based on Equation (9) associated with covariates  $\mathbf{x}_i = \{x_{ib_1}, x_{ib_2}, x_{ic_1}, x_{ic_2}\}$  and  $\mathbf{y}_i = \{y_{ib_1}, y_{ib_2}, y_{ic_1}, y_{ic_2}\}$ , respectively:

$$\begin{aligned} z_{ib_1}^* &= \alpha_{b_1} + \beta_{b_1}x_{ib_1} + \gamma_{b_1}y_{ib_1} + \varepsilon_{ib_1}, \\ z_{ib_2}^* &= \alpha_{b_2} + \beta_{b_2}x_{ib_2} + \gamma_{b_2}y_{ib_2} + \varepsilon_{ib_2}, \\ z_{ic_1} &= \alpha_{c_1} + \beta_{c_1}x_{ic_1} + \gamma_{c_1}y_{ic_1} + \varepsilon_{ic_1}, \\ z_{ic_2} &= \alpha_{c_2} + \beta_{c_2}x_{ic_2} + \gamma_{c_2}y_{ic_2} + \varepsilon_{ic_2}. \end{aligned} \tag{9}$$

We provided different simulation scenarios of the covariates values and three levels of correlation to analyze the response variables with the proposed model. The regression parameters were arbitrarily chosen and set to be fixed values in each simulation study. The errors in Equation (9) follow a multivariate normal distribution  $N_4(\mathbf{0}, \Sigma)$ , and the variance-covariance matrix  $\Sigma$  is given by

$$\begin{bmatrix} \sigma_{b_1}^2 & \sigma_{b_1}\sigma_{b_2}\rho_{b_1b_2} & \sigma_{b_1}\sigma_{c_1}\rho_{b_1c_1} & \sigma_{b_1}\sigma_{c_2}\rho_{b_1c_2} \\ \sigma_{b_1}\sigma_{b_2}\rho_{b_1b_2} & \sigma_{b_2}^2 & \sigma_{b_2}\sigma_{c_1}\rho_{b_2c_1} & \sigma_{b_2}\sigma_{c_2}\rho_{b_2c_2} \\ \sigma_{b_1}\sigma_{c_1}\rho_{b_1c_1} & \sigma_{b_2}\sigma_{c_1}\rho_{b_2c_1} & \sigma_{c_1}^2 & \sigma_{c_1}\sigma_{c_2}\rho_{c_1c_2} \\ \sigma_{b_1}\sigma_{c_2}\rho_{b_1c_2} & \sigma_{b_2}\sigma_{c_2}\rho_{b_2c_2} & \sigma_{c_1}\sigma_{c_2}\rho_{c_1c_2} & \sigma_{c_2}^2 \end{bmatrix}.$$

In the following data generating processes, the values of the variance-covariance parameters are set as  $\sigma_{b_1}^2 = 1$ ,  $\sigma_{b_2}^2 = 1$ ,  $\sigma_{c_1}^2 = 16$ , and  $\sigma_{c_2}^2 = 25$ , and the correlation is designed at the levels of low (all  $\rho = 0.3$ ), medium (all  $\rho = 0.5$ ), and high (all  $\rho = 0.7$ ), respectively, to assess the underlying model. Since there is no constraint on the sign of the correlation, the negative correlation can be estimated through our algorithm without further assumptions. Our simulation studies focus on the overall performance of the multivariate mixed response model through independent replications.

### 3.2.2. Point Estimates

Different simulation scenarios were designed to assess the performance of the underlying model on the point estimates by 1000 independent replications. There are two different sets of simulations for the data generating process, and within each setting, we analyze three levels of correlation, respectively. As shown in Table 1, the values of the regression parameters and the standard deviation of the continuous response variables are given across all simulation studies. In the first simulation setting, we provide 300 samples, and the mixed response variables are associated with covariates of distinct values. The covariate sets of  $\mathbf{x}_i$  and  $\mathbf{y}_i$  are identically and independently simulated from a normal distribution  $N(0, 1)$ , respectively, in each linear model.

In the second simulation setting, we present the multivariate mixed response model dealing with the common covariate. Regarding the data generating process with 1000 samples, all responses share one common covariate, which was generated from a normal distribution  $N(0, 1)$ , such as  $x_{ib_1} = x_{ib_2} = x_{ic_1} = x_{ic_2}$  in Equation (9). The second covariates  $\mathbf{y}_i$  are from a Bernoulli (0.5), and they are different

for each response. This setting represents the scenario in practice when a common factor is included in all of the response models.

**Table 1.** The ratio of the mean squared error (MSE) of the multivariate mixed response model (MMR) to the marginal model (GLM). Results based on 1000 independent simulation under two different scenarios and three different levels of correlation.

	Simulation I *			Simulation II †		
	Low	Med	High	Low	Med	High
$\alpha_{b_1} = 0.2$	1.00	0.99	0.98	0.97	0.92	0.85
$\beta_{b_1} = 0.3$	0.93	0.81	0.66	1.00	0.99	0.97
$\gamma_{b_1} = 0.3$	0.93	0.81	0.66	0.95	0.85	0.71
$\alpha_{b_2} = 0.2$	1.00	0.98	0.97	0.97	0.92	0.86
$\beta_{b_2} = 0.3$	0.94	0.84	0.69	1.00	0.98	0.95
$\gamma_{b_2} = 0.5$	0.94	0.83	0.70	0.95	0.86	0.71
$\alpha_{c_1} = 0.5$	1.00	1.00	1.00	0.96	0.89	0.79
$\beta_{c_1} = 8$	0.89	0.73	0.50	1.00	1.00	1.00
$\gamma_{c_1} = 10$	0.90	0.74	0.51	0.93	0.80	0.59
$\sigma_{c_1} = 4$	1.01	1.01	1.01	1.01	1.01	1.01
$\alpha_{c_2} = 0.4$	1.00	1.00	1.00	0.97	0.90	0.79
$\beta_{c_2} = 5$	0.92	0.77	0.53	1.00	1.00	1.00
$\gamma_{c_2} = 8$	0.92	0.75	0.50	0.94	0.80	0.57
$\sigma_{c_2} = 5$	1.01	1.01	1.01	1.01	1.01	1.01

\* Simulation I: N = 300, and the four responses have different covariates, e.g.,  $x_i$  and  $y_i$  generated from a normal distribution  $N(0, 1)$ ; † Simulation II: N = 1000, the responses shared one common covariate  $x_i \sim N(0, 1)$  and added a different covariate  $y_i \sim \text{Bernoulli}(0.5)$ .

In Table 1, we provide the ratio of the mean squared error (MSE) of the proposed method to the marginal approaches. This ratio represents the relative efficiency of the proposed method in comparison with the marginal method under different settings. In most of the simulation settings, the ratio rates of the MSE are well below 1. When the responses are highly correlated and have different covariate sets, our method can reduce MSE by 50%, which indicates a large efficiency gain.

### 3.2.3. Statistical Test

The test of composite-likelihood statistics can jointly assess the parameters of interest across different generalized linear models, while the conventional methods cannot achieve this. The simulation studies were conducted to measure the type I error rate and the power in comparison with the marginal approaches.

This simulation study was conducted to perform the hypothesis test. The correlated responses were generated based on Equation (9) with all correlation  $\rho = 0.3$ . The parameters of interest are the regression coefficients  $\{\beta_{b_1}, \beta_{b_2}, \beta_{c_1}, \beta_{c_2}\}$  of the first covariates across four generalized linear models, and the first covariates  $x_i$  are independently simulated from  $N(0, 1)$ . The regression coefficients of the second covariates  $y_i$  and other parameters are nuisance parameters with  $y_i \sim N(0, .5)$ . In the simulation study to assess the type 1 error rate, the regression parameters  $\{\beta_{b_1}, \beta_{b_2}, \beta_{c_1}, \beta_{c_2}\}$  are equal to zero in all generalized linear models, while other parameters have the same values as the previous simulation in Table 1. To assess the power, we fixed the values of the regression parameters as  $\beta_{b_1} = \beta_{b_2} = 0.1$  for the binary responses and  $\beta_{c_1} = \beta_{c_2} = 0.3$  for the continuous responses. In comparison, we combined the results of the marginal approaches by applying the Bonferroni adjustment.

Table 2 illustrates the results over 2000 independent replications. The proposed model analyzes all responses simultaneously, and the simulated type I error rates are valid and close to the 0.05. Through the test of the joint effect of the covariate of interest on all responses, the simulated power

is enhanced by our proposed model in comparison with the results from the Bonferroni test. As the sample size increases from 500 to 1000, the composite-likelihood statistics produce increased statistical power from approximately 0.800 to 0.989. The overall performance demonstrates that the composite statistics are more powerful than the conventional approach.

**Table 2.** Type 1 error rate and power under different sample sizes ( $N = 500$  and  $N = 1000$ ).

	Type I Error		Power	
	$N = 500$	$N = 1000$	$N = 500$	$N = 1000$
<b>Composite-Likelihood Method</b>				
$H_0 : \beta_{b_1} = \beta_{b_2} = \beta_{c_1} = \beta_{c_2} = 0$				
Likelihood ratio	0.054	0.043	0.804	0.988
Wald statistics	0.058	0.043	0.800	0.989
Scoring statistics	0.058	0.042	0.798	0.989
<b>Multiple Test</b>				
Bonferroni test	0.051	0.040	0.569	0.902

The likelihood ratio statistic is adjusted as Equation (7), which approximates to a  $\chi^2_4$ .

#### 4. Data Analysis

In this section, the multivariate mixed response model is applied to the clinical data from a colorectal cancer study. The data consist of clinical observations and demographic information on 743 patients, which are mixed with both categorical and continuous data. Our research interest was to evaluate the effect of treatment and other clinical factors on the toxicity outcomes. We focused on four common toxicity events that are related to colorectal cancer treatment. First, we chose nausea and diarrhea as two categorical responses. They are made up of ordinal data measuring the severity of the toxicity from grade 1 to 4. In our model setting, we only considered the occurrence of nausea and diarrhea for each patient. Therefore, these two responses were designed as binary variables, which are coded as 1 if they occurred and 0 if there is no record during the treatment. The continuous responses include two blood test measures, namely the hemoglobin (HGB) count and the white blood cell (WBC) count. Each patient had several blood examinations during the treatment, and we took the highest value for analysis. The explanatory variables contain the treatment effect, demographic information, tumor status, and genetic markers for each patient. There are two different treatment therapies in this colorectal cancer study. The patient demographic and clinical information, such as age, height, and weight, were collected as continuous variables. The tumor identified in either the colon or rectum was recorded as a binary variable. The study also includes the genetic markers, such as PERF1, PERF2, and KRAS, which are binary variables showing the occurrence of mutation. In total, we needed to jointly estimate 68 parameters for the coefficients of four linear models and the correlation between each outcome.

Table 3 shows the main result of the effect of the treatment, and the complete result is presented in Appendix B. We can observe that the statistical inference on the effect of the treatment through two models was in agreement. The second treatment therapy resulted in lower measures of the hemoglobin and indicates a negative association with the occurrence of nausea and diarrhea, whereas, the effect difference on the measures of white blood cells is insignificant. We can use the composite statistics to jointly assess the overall effect of this therapy on four responses. Table 4 provides the standard deviation and the correlation of the four clinical outcomes estimated based on the proposed model.

Using the conventional approach, we cannot make a statistic inference across different linear models. The proposed model is able to test the hypothesis  $H_0 : \beta_{b_1} = \beta_{b_2} = \beta_{c_1} = \beta_{c_2} = 0$  based on the asymptotical properties of the composite-likelihood function. The test statistics of the composite Wald statistics under the  $H_0$  is approximately 138.5890, the composite score statistics is 476.975, and the adjusted composite-likelihood ratio is 264.3069, which are all greater than the critical value of  $\chi^2_4$ . Therefore, we can reject the null hypothesis and conclude that the two different treatments

have a statistically significant difference in terms of patient toxicity response. More specifically, in our estimation results, we infer that there exists a significant difference in terms of the occurrence of nausea and diarrhea and a significant difference in HGB between the two treatments.

**Table 3.** The difference in the effect of treatment between the two treatment therapies. GLM: the generalized linear model; MMR: the multivariate mixed response model.

Regression Parameter	Models	
	GLM	MMR
<i>z<sub>b1</sub></i> : occurrence of nausea		
Intercept $\alpha_{b_1}$	$-0.2685 \pm 2.502$	$-0.2793 \pm 2.582$
( <i>p</i> value)	(0.833)	(0.832)
Treatment effect $\beta_{b_1}$	$-0.2644 \pm 0.190$	$-0.2724 \pm 0.193$
( <i>p</i> value)	(0.006)	(0.006)
<i>z<sub>b2</sub></i> : occurrence of diarrhea		
Intercept $\alpha_{b_2}$	$0.6631 \pm 2.557$	$0.6741 \pm 2.605$
( <i>p</i> value)	(0.611)	(0.612)
Treatment effect $\beta_{b_2}$	$-0.6231 \pm 0.192$	$-0.6422 \pm 0.198$
( <i>p</i> value)	(<0.001)	(<0.001)
<i>z<sub>c1</sub></i> : measures of hemoglobin		
Intercept $\alpha_{c_1}$	$160.3758 \pm 32.989$	$160.3775 \pm 32.984$
( <i>p</i> value)	(<0.001)	(<0.001)
Treatment effect $\beta_{c_1}$	$-12.492 \pm 2.498$	$-12.496 \pm 2.454$
( <i>p</i> value)	(<0.001)	(<0.001)
<i>z<sub>c2</sub></i> : measures of white blood cell		
Intercept $\alpha_{c_2}$	$12.295 \pm 9.331$	$12.2946 \pm 9.515$
( <i>p</i> value)	(0.010)	(0.011)
Treatment effect $\beta_{c_2}$	$-0.1591 \pm 0.706$	$-0.1597 \pm 0.702$
( <i>p</i> value)	(0.659)	(0.656)

**Table 4.** Estimation Results II: the estimated parameters contain second moments of each outcome.

	Estimated Correlation				Estimated Standard Deviation
	Nausea	Diarrhea	HGB	WBC	
Nausea	1.0000	0.3954	0.0736	0.0899	-
Diarrhea		1.0000	0.0351	-0.0126	-
HGB			1.0000	0.0139	16.796
WBC				1.0000	4.7507

### 5. Discussion

The problem of mixed outcomes is widely discussed in health-related studies. As a result of the computational complexity, most existing methods mainly focus on the case of two outcomes mixed with one continuous variable and one categorical variable. As an extension of the conditional grouped continuous model, we present the multivariate mixed response model to solve high-dimensional mixed multivariate regressions. Our model is constructed using the pairwise composite-likelihood method, such that multiple outcomes are analyzed through different bivariate models simultaneously. Regarding data mixed with continuous and binary responses, our method simplifies the problem of multiple outcomes into three types of scenario, which is both methodologically flexible and analytic appealing. From the simulation studies, the estimators of the proposed model demonstrate a lower MSE than the marginal approaches. The composite statistics also provide increased statistical power for joint hypothesis testing across different generalized linear models, which could make this a favorable approach to analyze clinical data with multiple mixed-type responses.

In addition, the model can be generalized to deal with ordinal and continuous data simultaneously. Under the same setup, the latent variable  $z_{ij}^*$  is normally distributed with mean  $\mu_{ij}^* = \mathbf{x}_{ij}^T \boldsymbol{\beta}_j$  and variance  $\sigma_j^2 = 1$ . If the latent normal variable  $z_{ij}^* \in [t_{l-1}, t_l)$  with the threshold values  $-\infty = t_0 < t_1 < \dots < t_l < \dots < t_{L-1} < t_L = \infty$ , the ordinal variable  $z_{ij} = l$  with  $l = 1, 2, \dots, L$ . Thus, the bivariate model of two ordinal outcomes can be formulated by the pairwise probability:

$$P(z_{ij} = l, z_{ik} = l') = P(z_{ij}^* \in [t_{l-1}, t_l), z_{ik}^* \in [t_{l'-1}, t_{l'})) = \int_{t_{l-1}}^{t_l} \int_{t_{l'-1}}^{t_{l'}} \phi_2(\omega_{ij}, \omega_{ik}; \mu_{ij}^*, \mu_{ik}^*, \rho_{jk}) d\omega_{ij} d\omega_{ik}.$$

The score function can be obtained by taking the derivatives of  $\sum_{i=1}^n \sum_{j=1}^{q-1} \sum_{k=j+1}^q \log P(z_{ij} = l, z_{ik} = l')$ , and the thresholds  $t_l$ 's need to be estimated with the monotone restriction. The maximum composite-likelihood estimation for mixed outcomes can be conducted through the same approaches as proposed in this paper. Further research will be considered to analyze multivariate outcomes with various distributions.

**Author Contributions:** Conceptualization, H.B. and X.G.; Methodology, H.B., Y.Z., and X.G.; Software, H.B. and Y.Z.; Validation, Y.Z., X.G. and W.X.; Formal Analysis, H.B., Y.Z., X.G., and W.X.; Investigation, Y.Z., X.G., and W.X.; Resources, W.X.; Data Curation, W.X.; Writing—Original Draft Preparation, H.B. and Y.Z.; Writing—Review & Editing, X.G. and W.X.; Supervision, X.G. and W.X.; Project Administration, X.G. and W.X.; Funding Acquisition, X.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** Zhong and Gao’s research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Derivatives for Mixed Response Variables

The score Function (6) contains the derivative of log marginal normal density and the log of the conditional probability:

$$U_{jk} = \frac{\partial}{\partial \theta_{jk}} \log P(\mathbf{z}_k) + \left[ \frac{\mathbf{1}(z_j = 0)}{\Phi(d_j)} - \frac{\mathbf{1}(z_j = 1)}{1 - \Phi(d_j)} \right] \frac{\partial \Phi(d_j)}{\partial \theta_{jk}}.$$

To illustrate the derivation, let  $\mathbf{z}_j$  and  $\mathbf{z}_k$  represent the  $n \times 1$  vector response variables with the design matrices  $X_j$  and  $X_k$ , respectively. Let

$$c_k = \frac{(\mathbf{z}_k - \boldsymbol{\mu}_k)}{\sigma_k}, \quad \text{and} \quad d_j = \frac{\mu_j^* + \rho_{jk} \frac{\sigma_j}{\sigma_k} (\mathbf{z}_k - \boldsymbol{\mu}_k)}{\sigma_j \sqrt{1 - \rho_{jk}^2}}.$$

The normal CDF has the following properties:

$$\Phi'(d_j) = \phi(d_j), \quad \phi'(c_k) = -c_k \phi(c_k), \quad \phi''(c_k) = (c_k^2 - 1) \phi(c_k).$$

The first component in the score function can be simplified as

$$\begin{aligned} \frac{\partial}{\partial \theta_{jk}} \log P(\mathbf{z}_k) &= \frac{\partial}{\partial \theta_{jk}} \log \left( \frac{1}{\sigma_k} \phi(c_k) \right) \\ &= \frac{1}{\phi(c_k)} \frac{\partial \phi(c_k)}{\partial \theta_{jk}} - \frac{1}{\sigma_k} \frac{\partial \sigma_k}{\partial \theta_{jk}} \\ &= -c_k \frac{\partial c_k}{\partial \theta_{jk}} - \frac{1}{\sigma_k} \frac{\partial \sigma_k}{\partial \theta_{jk}}, \\ \frac{\partial^2 \log P(\mathbf{z}_k)}{\partial \theta_{jk} \partial \theta_{jk}^T} &= \frac{1}{\sigma_k^2} \frac{\partial \sigma_k}{\partial \theta_{jk}} \frac{\partial \sigma_k}{\partial \theta_{jk}^T} - \frac{1}{\sigma_k} \frac{\partial^2 \sigma_k}{\partial \theta_{jk} \partial \theta_{jk}^T} - \frac{\partial c_k}{\partial \theta_{jk}} \frac{\partial c_k}{\partial \theta_{jk}^T} - c_k \frac{\partial^2 c_k}{\partial \theta_{jk} \partial \theta_{jk}^T}. \end{aligned}$$

The derivatives of the second component has a simple format with respect  $d_j$ :

$$\frac{\partial \Phi(d_j)}{\partial \theta_{jk}} = \phi(d_j) \frac{\partial d_j}{\partial \theta_{jk}},$$

$$\frac{\partial^2 \Phi(d_j)}{\partial \theta_{jk} \partial \theta_{jk}^T} = -d_j \phi(d_j) \frac{\partial d_j}{\partial \theta_{jk}} \frac{\partial d_j}{\partial \theta_{jk}^T} + \phi(d_j) \frac{\partial^2 d_j}{\partial \theta_{jk}^T \partial \theta_{jk}^T}.$$

By adding the results above, the score function and its derivative can be given by

$$\frac{\partial l_{jk}}{\partial \theta_{jk}} = -\frac{1}{\sigma_k} \frac{\partial \sigma_k}{\partial \theta_{jk}} - c_k \frac{\partial c_k}{\partial \theta_{jk}} + \left[ \frac{1(z_j = 0)}{\Phi(d_j)} - \frac{1(z_j = 1)}{1 - \Phi(d_j)} \right] \phi(d_j) \frac{\partial d_j}{\partial \theta_{jk}}$$

$$\frac{\partial^2 l_{jk}}{\partial \theta_{jk} \partial \theta_{jk}^T} = \left( \frac{1}{\sigma_k^2} \frac{\partial \sigma_k}{\partial \theta_{jk}} \frac{\partial \sigma_k}{\partial \theta_{jk}^T} - \frac{1}{\sigma_k} \frac{\partial^2 \sigma_k}{\partial \theta_{jk} \partial \theta_{jk}^T} - \frac{\partial c_k}{\partial \theta_{jk}} \frac{\partial c_k}{\partial \theta_{jk}^T} - c_k \frac{\partial^2 c_k}{\partial \theta_{jk} \partial \theta_{jk}^T} \right)$$

$$- \left[ \frac{1(z_j = 0)}{(\Phi(d_j))^2} + \frac{1(z_j = 1)}{(1 - \Phi(d_j))^2} \right] \left[ \phi(d_j)^2 \frac{\partial d_j}{\partial \theta_{jk}} \frac{\partial d_j}{\partial \theta_{jk}^T} \right]$$

$$+ \left[ \frac{1(z_j = 0)}{\Phi(d_j)} - \frac{1(z_j = 1)}{1 - \Phi(d_j)} \right] \times \left[ -d_j \phi(d_j) \frac{\partial d_j}{\partial \theta_{jk}} \frac{\partial d_j}{\partial \theta_{jk}^T} + \phi(d_j) \frac{\partial^2 d_j}{\partial \theta_{jk} \partial \theta_{jk}^T} \right].$$

Furthermore,

$$\frac{\partial c_k}{\partial \beta_k} = -\frac{1}{\sigma_k} X_k, \frac{\partial c_k}{\partial \sigma_k} = -\frac{1}{\sigma_k} c_k, \frac{\partial^2 c_k}{\partial \sigma_k^2} = \frac{2}{\sigma_k^2} c_k, \frac{\partial^2 c_k}{\partial \beta_k \partial \sigma_k} = \frac{1}{\sigma_k^2} X_k.$$

For the term  $d_j$ , more derivation is given here:

$$\frac{\partial d_j}{\partial \beta_k} = \frac{\rho_{jk} X_k}{\sigma_k \sqrt{1 - \rho_{jk}^2}},$$

$$\frac{\partial d_j}{\partial \beta_j} = -\frac{X_j}{\sigma_j \sqrt{1 - \rho_{jk}^2}},$$

$$\frac{\partial d_j}{\partial \sigma_k} = \frac{\rho_{jk} c_k}{\sigma_k \sqrt{1 - \rho_{jk}^2}},$$

$$\frac{\partial d_j}{\partial \sigma_j} = \frac{X_j \beta_j}{\sigma_j^2 \sqrt{1 - \rho_{jk}^2}},$$

$$\frac{\partial d_j}{\partial \rho_{jk}} = -\frac{c_k}{\sqrt{1 - \rho_{jk}^2}} - \frac{X_j \beta_j \rho_{jk}}{\sigma_j (1 - \rho_{jk}^2)^{\frac{3}{2}}} - \frac{\rho_{jk}^2 c_k}{(1 - \rho_{jk}^2)^{\frac{3}{2}}},$$

$$\frac{\partial^2 d_j}{\partial \sigma_k^2} = -\frac{2\rho_{jk} c_k}{\sigma_k^2 \sqrt{1 - \rho_{jk}^2}},$$

$$\begin{aligned} \frac{\partial^2 d_j}{\partial \sigma_j^2} &= -\frac{2X_j^T \beta_j}{\sigma_j^3 \sqrt{1 - \rho_{jk}^2}}, \\ \frac{\partial^2 d_j}{\partial \rho_{jk}^2} &= -\frac{c_k \rho_{jk}}{(1 - \rho_{jk}^2)^{\frac{3}{2}}} - \frac{X_j^T \beta_j}{\sigma_j} \left[ \frac{1}{(1 - \rho_{jk}^2)^{\frac{3}{2}}} + \frac{3\rho_{jk}^2}{(1 - \rho_{jk}^2)^{\frac{5}{2}}} \right] \\ &\quad - c_k \left[ \frac{2\rho_{jk}}{(1 - \rho_{jk}^2)^{\frac{3}{2}}} + \frac{3\rho_{jk}^3}{(1 - \rho_{jk}^2)^{\frac{5}{2}}} \right], \\ \frac{\partial^2 d_j}{\partial \rho_{jk} \partial \sigma_k} &= \frac{c_k}{\sigma_k} \left[ \frac{1}{\sqrt{1 - \rho_{jk}^2}} + \frac{\rho_{jk}^2}{(1 - \rho_{jk}^2)^{\frac{3}{2}}} \right], \\ \frac{\partial^2 d_j}{\partial \beta_j \partial \sigma_j} &= \frac{X_j}{\sigma_j^2 \sqrt{1 - \rho_{jk}^2}}, \\ \frac{\partial^2 d_j}{\partial \beta_j \partial \rho_{jk}} &= -\frac{X_j \rho_{jk}}{\sigma_j (1 - \rho_{jk}^2)^{\frac{3}{2}}}, \\ \frac{\partial^2 d_j}{\partial \beta_k \partial \rho_{jk}} &= \frac{X_k}{\sigma_k} \left[ \frac{1}{\sqrt{1 - \rho_{jk}^2}} + \frac{\rho_{jk}^2}{(1 - \rho_{jk}^2)^{\frac{3}{2}}} \right], \\ \frac{\partial^2 d_j}{\partial \sigma_k \partial \beta_k} &= -\frac{\rho_{jk} X_k}{\sigma_k^2 \sqrt{1 - \rho_{jk}^2}}. \end{aligned}$$

In conclusion, the score function  $U_{jk}(\theta_{jk})$  with respect to each parameter is given by

$$\begin{aligned} \frac{\partial l_{jk}}{\partial \beta_k} &= \frac{c_k}{\sigma_k} X_k + \left[ \frac{1(z_j = 0)}{\Phi(d_j)} - \frac{1(z_j = 1)}{1 - \Phi(d_j)} \right] \frac{\phi(d_j) \rho_{jk} X_k}{\sigma_k \sqrt{1 - \rho_{jk}^2}}, \\ \frac{\partial l_{jk}}{\partial \beta_j} &= \left[ \frac{1(z_j = 0)}{\Phi(d_j)} - \frac{1(z_j = 1)}{1 - \Phi(d_j)} \right] \left[ -\frac{\phi(d_j) X_j}{\sigma_j \sqrt{1 - \rho_{jk}^2}} \right], \\ \frac{\partial l_{jk}}{\partial \sigma_k} &= -\frac{1}{\sigma_k} + \frac{c_k^2}{\sigma_k} + \left[ \frac{1(z_j = 0)}{\Phi(d_j)} - \frac{1(z_j = 1)}{1 - \Phi(d_j)} \right] \frac{\phi(d_j) \rho_{jk} c_k}{\sigma_k \sqrt{1 - \rho_{jk}^2}}, \\ \frac{\partial l_{jk}}{\partial \sigma_j} &= \left[ \frac{1(z_j = 0)}{\Phi(d_j)} - \frac{1(z_j = 1)}{1 - \Phi(d_j)} \right] \phi(d_j) \frac{X_j \beta_j}{\sigma_j^2 \sqrt{1 - \rho_{jk}^2}}, \\ \frac{\partial l_{jk}}{\partial \rho_{jk}} &= \left[ \frac{1(z_j = 0)}{\Phi(d_j)} - \frac{1(z_j = 1)}{1 - \Phi(d_j)} \right] \phi(d_j) \\ &\quad \times \left[ -\frac{c_k}{\sqrt{1 - \rho_{jk}^2}} - \frac{X_j \beta_j \rho_{jk}}{\sigma_j (1 - \rho_{jk}^2)^{\frac{3}{2}}} - \frac{\rho_{jk}^2 c_k}{(1 - \rho_{jk}^2)^{\frac{3}{2}}} \right]. \end{aligned}$$

**Appendix B. Estimation Results of the Clinical Data From a Colorectal Cancer Study**

**Table A1.** The estimated regression coefficients  $\beta$  and the standard deviation (sd). GLM: the estimation via the generalized linear model; MRM: estimation via proposed multivariate mixed response model; the column of \* lists the significant covariates.

Parameters	1. Nausea		2. Diarrhea		3. HBG		4. WBC		*
	GLM	MMR	GLM	MMR	GLM	MMR	GLM	MMR	
Intercept	-0.2685 (1.277)	-0.2793 (1.317)	0.6631 (1.304)	0.6741 (1.329)	160.3759 (16.831)	160.3775 (16.829)	12.2949 (4.761)	12.2946 (4.855)	3, 4
Treatment	-0.2644 (0.097)	-0.2724 (0.099)	-0.6231 (0.098)	-0.6422 (0.101)	-12.4921 (1.274)	-12.4957 (1.252)	-0.1591 (0.360)	-0.1597 (0.358)	1, 2, 3
OS	-0.0037 (0.009)	-0.0039 (0.009)	0.0219 (0.009)	0.0228 (0.009)	0.5674 (0.115)	0.5675 (0.117)	-0.1764 (0.032)	-0.1764 (0.029)	2, 3, 4
OS event	0.0650 (0.195)	0.0702 (0.194)	0.1173(0.200)	0.1208 (0.194)	1.5154 (2.592)	1.5159 (2.614)	0.1290 (0.733)	0.1295 (0.562)	
PFS	0.0305 (0.020)	0.0315 (0.020)	0.0066 (0.021)	0.0062 (0.022)	-0.1628 (0.259)	-0.1627 (0.258)	0.1242 (0.073)	0.1243 (0.062)	
IN	-0.0204 (0.188)	-0.0200 (0.190)	-0.8622 (0.203)	-0.8907 (0.212)	-11.1599 (2.486)	-11.1621 (2.932)	-0.0809 (0.703)	-0.0803 (0.810)	2, 3
PD	0.1263 (0.136)	0.1319 (0.139)	-0.1375 (0.136)	-0.1432 (0.142)	-7.1013 (1.782)	-7.1005 (1.667)	-0.3695 (0.504)	-0.3689 (0.528)	3
PR	-0.0213 (0.169)	-0.0207 (0.171)	0.0691 (0.176)	0.0707 (0.174)	3.6782 (2.234)	3.6781 (2.210)	0.4003 (0.632)	0.4005 (0.5277)	
Age	-0.0142 (0.008)	-0.0091 (0.005)	0.0029 (0.008)	0.0018 (0.005)	-0.1463 (0.061)	-0.1464 (0.061)	-0.0382 (0.017)	-0.0382 (0.018)	3, 4
Gender	-0.0086 (0.005)	-0.0091 (0.005)	0.0016 (0.005)	0.0018 (0.005)	9.0894 (1.798)	9.0836 (1.825)	-0.1079 (0.509)	-0.1096 (0.534)	1, 3
Colon	0.2313 (0.165)	0.2378 (0.169)	0.1017 (0.170)	0.1059 (0.177)	-1.1366 (2.189)	-1.1338 (2.105)	0.5706 (0.619)	0.5713 (0.676)	
Rectum	-0.0075 (0.154)	-0.0111 (0.157)	0.1760 (0.158)	0.1844 (0.166)	-1.6947 (2.032)	-1.6940 (1.913)	0.9096 (0.575)	0.9094 (0.649)	
Height	0.0048 (0.007)	0.0050 (0.008)	-0.0050 (0.007)	-0.0051 (0.008)	-0.0820 (0.097)	-0.0820 (0.097)	0.0237 (0.028)	0.0237 (0.028)	
Weight	0.0023 (0.003)	0.0024 (0.003)	0.0012 (0.003)	0.0012 (0.003)	0.0906 (0.044)	0.0907 (0.042)	-0.0189 (0.012)	-0.0189 (0.012)	3
PERF1	0.2046 (0.107)	0.2132 (0.110)	0.0898 (0.109)	0.0933 (0.110)	-4.8490 (1.414)	-4.8469 (1.352)	0.6028 (0.400)	0.6036 (0.380)	3
PERF2	0.5041 (0.187)	0.5244 (0.187)	0.2637 (0.189)	0.2692 (0.193)	-7.1494 (3.814)	-7.1455 (2.628)	1.8158 (0.688)	1.8176 (0.685)	1, 3, 4
KRAS	-0.0407 (0.292)	-0.0387 (0.295)	0.0080 (0.293)	0.0150 (0.294)	1.3456 (3.814)	1.3455 (3.285)	-0.7323 (1.079)	-0.7325 (0.936)	

In Section 4, we illustrate some estimation results of the regression parameters and correlation between outcomes. The complete analysis of these data consists of 68 parameters.

## References

1. Cox, D.R. The analysis of multivariate binary data. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1972**, *21*, 113–120. [[CrossRef](#)]
2. Cox, D.R.; Wermuth, N. Response models for mixed binary and quantitative variable. *Biometrika* **1992**, *79*, 441–461. [[CrossRef](#)]
3. Olkin, L.; Tate, R.F. Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Stat.* **1961**, *32*, 448–456. [[CrossRef](#)]
4. Teixeira-Pinto, A.; Normand, S.T. Correlated bivariate continuous and binary outcomes: Issues and applications. *Stat. Med.* **2009**, *28*, 1753–1773. [[CrossRef](#)] [[PubMed](#)]
5. Sammel, M.D.; Ryan, L.M.; Legler, J.M. Latent variable models for mixed discrete and continuous outcomes. *J. R. Stat. Soc. Ser. B Methodol.* **1997**, *59*, 667–678. [[CrossRef](#)]
6. Sammel, M.D.; Lin, X.; Ryan, L.M. Multivariate linear mixed models for multiple outcomes. *Stat. Med.* **1999**, *18*, 2479–2492. [[CrossRef](#)]
7. Yang, Y.; Kang, J.; Mao, K.; Zhang, J. Regression models for mixed poisson and continuous longitudinal data. *Stat. Med.* **1961**, *26*, 3782–3800. [[CrossRef](#)]
8. Anderson, J.A.; Pemberton, J.D. The grouped continuous model for multivariate ordered categorical variables and covariate adjustment. *Biometrics* **1985**, *41*, 875–885. [[CrossRef](#)]
9. De Leon, A. Pairwise likelihood approach to grouped continuous model and its extension. *Stat. Probab. Lett.* **2005**, *75*, 49–57. [[CrossRef](#)]
10. Poon, W.Y.; Lee, S.Y. Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika* **1987**, *52*, 409–430. [[CrossRef](#)]
11. Skrondal, A.; Rabe-Hesketh, S. *Latent Variable Modelling: A Survey*; Blackwell: Oxford, UK, 2007; Volume 34.
12. Catalano, P.; Ryan, L. Bivariate latent variable models for clustered discrete and continuous outcomes. *J. Am. Stat. Assoc.* **1992**, *50*, 1078–1095. [[CrossRef](#)]
13. Catalano, P.J. Bivariate modeling of clustered continuous and ordered categorical outcomes. *Stat. Med.* **1997**, *16*, 883–900. [[CrossRef](#)]
14. Najita, J.S.; Li, Y.; Catalano, P.J. A novel application of a bivariate regression model for binary and continuous outcomes to studies of fetal toxicity. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2009**, *58*, 555–573. [[CrossRef](#)] [[PubMed](#)]
15. Gueorguieva, R.V.; Agresti, A. A correlated probit model for joint modeling of clustered binary and continuous responses. *J. Am. Stat. Assoc.* **2001**, *96*, 1102–1112. [[CrossRef](#)]
16. Zhang, H.; Liu, D.; Zhao, J.; Bi, X. Modeling hybrid traits for comorbidity and genetic studies of alcohol and nicotine co-dependence. *Ann. Appl. Stat.* **2018**, *12*, 2359–2378. [[CrossRef](#)]
17. De Leon, A.R.; Carriégre, K.C. General mixed-data model: Extension of general location and grouped continuous models. *Can. J. Stat.* **2007**, *35*, 533–548. [[CrossRef](#)]
18. De Leon, A.R.; Carriégre, K.C. *Analysis of Mixed Data*; Chapman and Hall/CRC: New York, NY, USA, 2013.
19. Lindsay, B. Composite likelihood methods. *Contemp. Math.* **1988**, *80*, 220–239.
20. Cox, D.R.; Reid, N. A note on pseudolikelihood constructed from marginal densities. *Biometrika* **2004**, *91*, 729–737. [[CrossRef](#)]
21. Varin, C. On composite marginal likelihoods. *Adv. Stat. Anal.* **2008**, *92*, 1–28. [[CrossRef](#)]
22. Varin, C.; Reid, N.; Firth, D. An overview of composite likelihood methods. *Stat. Sin.* **2011**, *21*, 5–42.
23. Reid, N.; Lindsay, B.; Liang, K.Y. Composite likelihood methods introduction. *Stat. Sin.* **2011**, *21*, 1–3.
24. Faes, C.; Aerts, M.; Molenberghs, G.; Geys, H.; Teuns, G.; Bijnen, L. A high-dimensional joint model for longitudinal outcomes of different nature. *Stat. Med.* **2008**, *27*, 4408–4427. [[CrossRef](#)] [[PubMed](#)]
25. Atkinson, K.E. *An Introduction to Numerical Analysis*; John Wiley and Sons Inc.: New York, NY, USA, 1989.
26. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; Prentice Hall: Upper Saddle River, NJ, USA, 1998.
27. Dunson, D.B. Bayesian latent variable models for clustered mixed outcomes. *J. R. Stat. Soc.* **2000**, *62*, 355–366. [[CrossRef](#)]
28. Cox, D.R.; Hinkley, D.V. *Theoretical Statistics*; Chapman and Hall: London, UK, 1974.
29. Kent, J.T. Robust properties of likelihood ratio tests. *Biometrika* **1982**, *69*, 19–27.
30. Jin, Z. Aspects of Composite Likelihood Inference. Ph.D. Thesis, University of Toronto, Toronto, ON, Canada, 2010.

31. Gao, X.; Song, P.X.K. Composite likelihood bayesian information criteria for model selection in high-dimensional data. *J. Am. Stat. Assoc.* **2010**, *105*, 1531–1540. [[CrossRef](#)]
32. Godambe, V. An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.* **1960**, *31*, 1208–1211. [[CrossRef](#)]
33. Rotnitzky, A.; Jewell, N.P. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **1990**, *77*, 485–497. [[CrossRef](#)]
34. Geys, H.; Molenberghs, G.; Ryan, LM. Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *J. Am. Stat. Assoc.* **1999**, *94*, 734–745. [[CrossRef](#)]
35. Pace, L.; Salvan, A.; Sartori, N. Adjusting composite likelihood ratio statistics. *Stat. Sin.* **2011**, *21*, 129–148.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).