Mahdis Azadbakhsh[1] / Xin Gao[1] / Hanna Jankowski[1]

Multiple Comparisons Using Composite Likelihood in Clustered Data

[1] Department of Statistics and Mathematics, York University, Toronto, ON M3J 1P3, Canada, E-mail: xingao@mathstat.yorku.ca

Abstract:
We study the problem of multiple hypothesis testing for correlated clustered data. As the existing multiple comparison procedures based on maximum likelihood estimation could be computationally intensive, we propose to construct multiple comparison procedures based on composite likelihood method. The new test statistics account for the correlation structure within the clusters and are computationally convenient to compute. Simulation studies show that the composite likelihood based procedures maintain good control of the familywise type I error rate in the presence of intra-cluster correlation, whereas ignoring the correlation leads to erratic performance.

Keywords: multiple comparisons, composite likelihood, strong control of type I error

DOI: 10.1515/ijb-2016-0004

1 Introduction

The prevalence of depression in seniors estimated by the World Health Organization varies between 10% to 20% [1]. Understanding the relationship between depression and other health factors can help prevent the disease and alleviate the symptoms. The health and retirement study (HRS) conducted by the University of Michigan is a longitudinal study measuring various aspects of health, retirement and aging, as well as depression status. In this study, seniors were measured every two years from 1994 to 2012. The objective of our analysis is to estimate the effect of several health factors known to be associated with depression status and compare the effect sizes of different factors. Multiple comparisons on the effect sizes will clarify the relative importance of factors on the disease. For example, sleeplessness and smoking both contribute to the occurrence rate of depression. One might question if one factor is more important than the other for development of the disease. Therefore, to fully understand the effects of the health factors, we perform multiple comparisons on their effect sizes. It is important to note that we only observe associations but do not establish causal relationships of health factors with the disease.

The repeated binary measurements of depression status observed in this data set are correlated within individuals. These repeated measurements can be viewed as "clustered" data since they are recorded from the same experimental unit multiple times. Examples of clustered data arise in many other situations, including measurements coming from siblings or same pedigrees, or measurements taken in close proximity to each other in spatial data. Ignoring existing correlations within clusters leads to invalid individual or multiple inferences.

When performing multiple comparisons in clustered data, one should therefore take into account the correlation structure within the clusters. However, full likelihood analyses on such data often encounter computational challenges. Composite likelihood methods are extensions of the likelihood method that project high-dimensional likelihood functions to low-dimensional ones [2, 3]. This dimension reduction is achieved by compounding valid marginal or conditional densities. It has been shown that, under regularity conditions, the composite likelihood estimator has desirable properties, such as consistency and asymptotic normality [2–5]. This makes it an appealing alternative in inferential procedures. Furthermore, composite likelihood is more computationally convenient than full likelihood at a cost of some loss of efficiency. The magnitude of this loss depends on the dimension of the multivariate vector and its dependency structure. Composite likelihood methodology has been applied to numerous statistical problems [6–8], however, the potential of composite likelihood in multiple testing has yet to be explored.

Multiple testing procedures have been developed to control the overall type I error rate when the number of tests is greater than one [9, 10]. Family-wise error rate (FWER) is defined as the probability of falsely rejecting at least one true null hypothesis. The classical Bonferroni method is the simplest procedure to adjust the FWER. It assumes maximum negative correlation between test statistics, and the resulting FWER does not exceed the significance level in any setting given the per comparison error is controlled correctly. However, it is well-known

Xin Gao is the corresponding author.

to be very conservative. The Dunn-Sidák procedure [11] uses a slightly less conservative p-value threshold for each comparison. This procedure assumes uncorrelated test statistics that explains the power increase compared to the Bonferroni method. The procedure provides the exact control of the FWER under independence, and it is conservative under positive dependence and it is liberal under negative dependence. For example, testing T_i and $-T_i$ simultaneously is not possible with Sidák. Schéffe [12] established a method for testing all possible linear comparisons among a set of normally distributed variables, which tends to be over-conservative for a finite family of multiple comparisons. Scheffé controls all possible linear combinations, which makes a comparison with the other methods quite difficult, as they are testing different sets of hypotheses, where a smaller set of hypotheses of interest will in most cases result in a higher power. Several stage-wise procedures have also been proposed to improve the power. All of these methods present some shortcuts to avoid testing all intersection hypotheses in closed testing. Simes [13] modified the Bonferroni procedure based on ordered p-values. Holm [14] proposed a multi-stage procedure based on closed testing procedure. The method adjusts the FWER in each step using the number of remaining null hypotheses. Hommel [15] suggested a stage-wise rejective multiple test based on Simes inequality. All of these methods are less conservative and therefore more powerful than the Bonferroni method. However, it is difficult to construct simultaneous confidence intervals based on a stage-wise approach. As another alternative, Hothorn et al. [16] proposed to use equi-coordinate quantiles of the multivariate normal and multivariate t distribution to perform multiple comparisons in parametric methods. This corresponds to the control of the FWER based on the maximum of test statistics. The MNQ method allows for any correlation between test statistics, but needs the assumption of an asymptotic multivariate normal distribution for the test statistics, which is most of the time validated by the multivariate central limit theorem. This method offers sharp control of the FWER. The approach has been employed in many parametric and nonparametric settings to provide both multiple inferences and simultaneous intervals [17, 18].

In this paper, we propose a new procedure where multiple testing methods are combined with composite likelihood inference. This is motivated by the multiple testing problems for which full likelihood inference can be rather computational complex. We show that the composite likelihood test statistics for multiple hypotheses offer strong control of the FWER with the use of closed testing procedures. We explore in detail different multivariate models for correlated clustered data including the multivariate normal, multivariate probit, and quadratic exponential models to illustrate our multiple comparisons approach. Among these methods, the multivariate normal quantile threshold appears to have the best control of the FWER in most simulation settings.

The structure of this paper is as follows: In Section 2, we develop our composite likelihood based test statistics for multiple inferences and establish their asymptotic properties. In Section 3, we provide details on how to apply the general approach for several multivariate models. In Section 4, we conduct simulation studies to evaluate empirical performance of the proposed method. Finally, we analyze the depression data set to demonstrate the practical utility of the method.

2 Multiple comparisons procedures based on composite likelihood

Let $\{f(Y;\theta), \theta \in \Theta\}$, where $\theta = (\theta_1, \ldots, \theta_p)^T$, be a parametric statistical model with parameter space $\Theta \subset \mathbb{R}^p$. Let $Y = (y_1^T, \cdots, y_n^T)$ denote the response variables, where $y_i = (y_{i1}, \cdots, y_{im_i})^T$ is the vector of observations from cluster i, $i = 1, \cdots, n$ from a study population. It is assumed that observations across different clusters are independent, whereas observations within the same cluster may be dependent. Note that overall sample size is $\sum_{i=1}^{n} m_i$. We assume that the cluster size, m_i, is uniformly bounded.

Let

$$C = C_{c\times p} = (C^{(1)}, C^{(2)}, \cdots, C^{(c)})^T$$

denote the contrast matrix. A family of c linear combinations of the parameters can then be specified by $C\theta$. Let H_{0l} denote the hypothesis that $C^{(l)}\theta = 0$, for $l = 1, \ldots, c$. We focus here on jointly testing the family of hypotheses $H_{0l}, l = 1, \ldots, c$. The multiple testing procedure has weak control of the FWER if the FWER $\leq \alpha$ when all of the null hypotheses are true, whereas one has strong control of the FWER if the FWER $\leq \alpha$ under any combinations of null hypotheses and alternative hypotheses. Composite likelihood was proposed as a pseudo likelihood inference method that has attracted much attention in recent years [3–5, 19]. Two popular examples of composite likelihood are the "univariate marginal likelihood" and "univariate conditional likelihood". In the former, we have the composite likelihood function$CL(\theta; Y) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} f(y_{ij})$,

where any dependence structure is ignored. In the latter, we take $CL(\theta; Y) = \prod_{i,j} f(y_{ij}|y_{i(-j)}) = \prod_{i,j} f(y_i)/f(y_{i(-j)})$, where $y_{i(-j)}$ denotes the sub-vector of y_i with its j th element removed. We consider both types of approaches in our examples. In general, composite likelihood is a compounded form of marginal or conditional likelihoods, which is often easier to maximize than full likelihood. For $A_k \subset \{(i,j) : j = 1, \dots, m_i, i = 1, \dots, n\}$, let $Y_{A_k} = \{Y_{ij}, (i,j) \in A_k\}$ denote a subset of the data, where $k = 1, \dots, K$. The composite likelihood function is then defined as

$$CL(\theta; Y) = \prod_{k=1}^{K} f(y_{A_k}; \theta)^{w_{A_k}},$$

where $f(y_{A_k}; \theta)$ is the density for the subset vector y_{A_k}, and w_{A_k} are some suitably chosen weights. In practice, the type of composite likelihood should be chosen so that the resulting composite score equation is consistent for the parameters, and the computation complexity is sufficiently manageable.

The maximum composite likelihood estimate (MCLE) is defined as

$$\hat{\theta}_n^c = \text{argmax}_{\theta \in \Theta} CL(\theta; Y).$$

Xu and Reid [20] give precise conditions under which $\hat{\theta}_n^c$ is consistent for θ. Under appropriate assumptions, $\sqrt{n}(\hat{\theta}_n^c - \theta)$ is also asymptotically normally distributed with mean zero and limiting variance given by the inverse of the the Godambe information matrix [3, 21], where

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

with $H(\theta) = \lim_n E(-cl^{(2)}(\theta; Y))/n$ and $J(\theta) = \lim_n \text{var}(cl^{(1)}(\theta; Y))/n$. Here, $cl^{(1)}$ is the vector of first derivatives and $cl^{(2)}$ is the matrix of second order derivatives of $cl(\theta; Y) = \log CL(\theta; Y)$ with respect to θ. The matrix $H(\theta)$ can be estimated as the negative Hessian matrix evaluated at the maximum composite likelihood estimator, whereas the matrix $J(\theta)$ can be estimated as the sample covariance matrix of the composite score vectors. Both estimators, which we denote as $\widehat{H}_n$ and $\hat{J}_n$, are consistent [21].

Consider the hypothesis test on a family of linear combinations of the parameters: $\{H_0 : C\theta = 0\}$. Denote by $\Gamma = G^{-1}(\theta)$ the inverse Godambe information matrix, and let $\hat{\Gamma}_n$ denote the consistent estimator of Γ, where $\hat{\Gamma}_n = \widehat{H}_n^{-1} \hat{J}_n \widehat{H}_n^{-1}$. We propose the following test statistics for our hypothesis test

$$T_{l,n} = \frac{C^{(l)T} \hat{\theta}_n^c}{\sqrt{\left(C^{(l)T} \hat{\Gamma}_n C^{(l)}\right)/n}}, \quad l = 1, \dots, c. \tag{2}$$

The limiting distribution of $T_n = (T_{1,n}, \cdots, T_{l,n})^T$ is multivariate normal $MVN(0, V)$, where

$$V = \text{diag}(D)^{-1/2} D \, \text{diag}(D)^{-1/2}, \quad D = CG^{-1}(\theta)C^T. \tag{3}$$

Furthermore, since $V_{i,i} = 1$, the marginal asymptotic distribution of each individual $T_{l,n}$ is standard normal. In practice, we estimate V by plugging $\hat{\Gamma}_n$ as a consistent estimator of $G^{-1}(\theta)$ into eq. (3). This results in a consistent estimator of V.

We illustrate a few multiple comparison procedures as follows.

1. The Bonferroni procedure: The global intersection hypothesis $\cap_{l=1}^c H_{0l}$ will be rejected if $\max_l |T_{l,n}| > Z_{1-(\alpha/2c)}$, where $Z_{1-(\alpha/2c)}$ denotes the critical value for standard normal random variable with $P(Z > Z_{1-(\alpha/2c)}) = \alpha/2c$. Each individual hypothesis H_{0l} will be rejected if $|T_{l,n}| > Z_{1-(\alpha/2c)}$.

2. The Holm's procedure: For each H_{0l}, evaluate the p-value $p_l = 2P(Z > |T_{l,n}|)$. Order the p-values from the least to the greatest as $p_{(1)}, \dots, p_{(c)}$ and the corresponding hypotheses are reordered as $H_{(01)}, \dots, H_{(0c)}$. The global intersection hypothesis $\cap_{l=1}^c H_{0l}$ will be rejected if $p_{(1)} \leq \alpha/c$. Let k denote the smallest l so that $p_{(l)} > \alpha/(c - l + 1)$. If $k > 1$, then the individual hypotheses $H_{(01)}, \dots, H_{(0,k-1)}$ will be rejected.

3. The MNQ procedure: The global intersection hypothesis $\cap_{l=1}^c H_{0l}$ will be rejected if $\max_l |T_{l,n}| > Q_{1-\alpha,V}$, where $Q_{1-\alpha,V}$ denotes the equi-coordinate quantile for multivariate normal random vector $Z = (Z_1, \dots, Z_c)$ with $Z \sim N(0, V)$, and $P(\max_l |Z_l| > Q_{1-\alpha,V}) = \alpha$. Each individual hypothesis H_{0l} will be rejected if $|T_{l,n}| > Q_{1-\alpha,V}$.

As the variance estimators are consistent regardless of the configuration of the hypotheses, the multivariate distribution of any subset of the test statistics still follows a multivariate normal distribution. Therefore, if a closed testing procedure is applied on the test statistics, the procedure has a strong control of the FWER [22, 23].

It is worthy to point out that the test statistics we propose here are Wald-type statistics which are not invariant under re-parametrization. Under re-parametrization, the new statistics follow the same type of limiting distributions, but the values of the statistics are not the same. This is a standard limitation associated with Wald-type statistics.

The multivariate distribution of T_n can be approximated by a multivariate t distribution. The denominator ${C^{(l)}}^T\hat{\Gamma}_n C^{(l)}/n$ has an asymptotic equivalent distribution as ${C^{(l)}}^T H^{-1}\hat{J}_n H^{-1}C^{(l)}/n$ based on Slutsky' Theorem. Furthermore, $\hat{J}_n = (cl^{(1)})^T cl^{(1)}$ asymptotically follows a Wishart (J, n). This entails that asymptotically ${C^{(l)}}^T H^{-1}\hat{J}_n H^{-1}C^{(l)}$ follows $\sigma_l^2\chi_n^2$, where $\sigma_l^2 = {C^{(l)}}^T H^{-1}JH^{-1}C^{(l)}$. Reformulate the test statistics as

$$T_{l,n} = \frac{{C^{(l)}}^T\hat{\theta}_n^c/\sigma_l}{\sqrt{\left({C^{(l)}}^T\hat{\Gamma}_n C^{(l)}\right)/(n\sigma_l^2)}},$$

where the numerator is asymptotically a multivariate normal $MVN\ (0, V)$, and the denominator is asymptotically $\sqrt{\chi_n^2/n}$. Therefore, the multivariate distribution of T_n can be approximated as a multivariate $t(V, n)$, where V is the covariance matrix and n is the degrees of freedom.

The MNQ method also facilitates the construction of simultaneous confidence intervals. The simultaneous $(1-\alpha)100\%$ confidence interval for $C\theta$ is

$$\left({C^{(l)}}^T\hat{\theta}_n^c \pm Q_{\alpha,V}\sqrt{\left({C^{(l)}}^T\hat{\Gamma}_n C^{(l)}\right)/n}\right). \tag{4}$$

In some applications, the collection of effect sizes are nonlinear monotone transformations of the parameters. For example, we obtain odds ratio from log odds ratio by applying the exponential function. Let $G(.)$ denote a nonlinear monotone transformation. Then the simultaneous $(1-\alpha)100\%$ confidence interval for $G[(C^{(l)})^T\theta]$, $l = 1, \ldots, c$, is

$$\left(G\left[{C^{(l)}}^T\hat{\theta}_n^c \pm Q_{\alpha,V}\sqrt{\left({C^{(l)}}^T\hat{\Gamma}_n C^{(l)}\right)/n}\right]\right). \tag{5}$$

3 Three multivariate models

To demonstrate the application of our methodology, we consider three specific multivariate distributions: multivariate normal, multivariate probit, and quadratic exponential distributions. In this Section, we give the details of the distribution and computation of the test statistics; Section 4 examines the behavior of the test statistics via simulations.

For the first two distributions, the composite likelihood is constructed as a product of univariate marginal likelihoods. We choose this type of composite likelihood as the univariate marginal likelihood of these two distributions follows the univariate exponential family for which the estimation of parameters is not difficult. For the quadratic exponential distribution, the composite likelihood is constructed as a conditional likelihood. For the quadratic exponential model the normalizing constant is computationally intensive to obtain, and in this composite likelihood the normalizing constant is canceled out, greatly simplifying the computational burden of estimation.

In order to include covariates into our modelling scheme, let X_i denote an $m_i \times p$ matrix containing the values of p covariates for the m_i individuals in the i^{th} cluster and $\beta = (\beta_1, \ldots, \beta_p)^T$ denote the vector of regression coefficients. Let $\vec{x}_{ij}$ denote the j^{th} row of the matrix X_i (this is the vector of covariates for individual j in cluster i).

3.1 Multivariate Gaussian distribution

Let $\{(y_i, X_i),\ i = 1, \cdots n\}$, denote the response and covariates arising from a multivariate normal model, with$y_i = X_i\beta + \epsilon_i, i = 1, \ldots, n$, and $m_i = m$. We assume that $\epsilon_i \sim N_m(0, \Sigma)$ where $\Sigma = (\sigma_{ij}), i, j = 1, \ldots, m$, is an arbitrary covariance matrix. The univariate composite likelihood is thus equal to

$$cl\,(\beta) \;=\; \textstyle\sum_{i=1}^{n}\sum_{j=1}^{m}(-\frac{1}{2}\log(2\pi\sigma_{jj}) - \frac{1}{2\sigma_{jj}^{2}}(y_{ij} - \vec{x}_{ij}\beta)^{2}), \qquad (6)$$

where the σ_{jj}'s are nuisance parameters. The Hessian matrix and variability matrix are, respectively, $H(\beta) = n^{-1}\left(\sum_{i=1}^{n} X_i^T W X_i\right)$ and $J(\beta) = n^{-1}\left(\sum_{i=1}^{n} X_i^T W \Sigma W X_i\right)$, with $W = \text{diag}(\Sigma)^{-1}$. To estimate the regression coefficients, we employ an iterative algorithm: Given the current estimate for the nuisance parameters σ_{jj}'s, we maximize the composite likelihood to obtain an estimate of $\widehat{\beta}_n^c = (\sum_{i=1}^{n} X_i^T W X_i)^{-1}\sum_{i=1}^{n} X_i^T W Y_i$, and given a current estimate for β, we use the sample covariance matrix of residuals to estimate Σ. Based on the estimates $\widehat{\beta}_n^c$ and $\widehat{\Sigma}$, we obtain estimates for $H(\beta)$ and $J(\beta)$ with W being replaced by its estimate $\widehat{W} = \text{diag}(\widehat{\Sigma})$.

3.2 Multivariate probit model

Let $y_i^* = X_i\beta + \epsilon_i$ with $\epsilon_i \sim N_m(0, \Sigma)$ and $\Sigma = \sigma R$, where R is an $m \times m$ correlation matrix. The variables y_i^* are the latent response variables, and their dichotomized version of the latent variable with $y_{ij} = I(y_{ij}^* > 0),\ j = 1, \cdots, m$ yield the multivariate probit model. We therefore have that $P(y_{ij} = 1|X_i) \;=\; \Phi(\vec{x}_{ij}\beta/\sigma)$ where Φ denotes the univariate standard normal cumulative distribution function. It follows that the parameters β and σ are not fully identifiable in the model, and we can only estimate the ratio β/σ. To simplify notation, σ is set equal to 1 in what follows. The univariate composite log-likelihood function of the probit model is then formulated as

$$cl(\beta; Y) \;=\; \textstyle\sum_{i=1}^{n}\sum_{j=1}^{m}[y_{ij}\ \log\ \Phi\left(\vec{x}_{ij}\beta\right) + (1 - y_{ij})\ \log\left(1 - \Phi\left(\vec{x}_{ij}\beta\right)\right)].$$

Denoting $\mu_{ij} = P(y_{ij} = 1|X_i)$, and $\mu_i = (\mu_{i1}, \ldots, \mu_{im})^T$, we have

$$cl^{(1)}(\beta; Y) \;=\; \textstyle\sum_{i=1}^{n}\left(\frac{\partial\mu_i}{\partial_\beta}\right)^T \Pi_i^{-1}(y_i - \mu_i),$$

where $\Pi_i = \text{diag}(\text{var}(y_{i1}), \cdots, \text{var}(y_{im}))$, and $\text{var}(y_{ij}) = \mu_{ij}(1 - \mu_{ij})$. This yields

$$\begin{aligned} H(\beta) &= n^{-1}\sum_{i=1}^{n}\left(\frac{\partial\mu_i}{\partial_\beta}\right)^T \Pi_i^{-1}\left(\frac{\partial\mu_i}{\partial_\beta}\right), \text{ and} \\ J(\beta) &= n^{-1}\sum_{i=1}^{n}\left(\frac{\partial\mu_i}{\partial_\beta}\right)^T \Pi_i^{-1}\,\text{cov}(y_i)\,\Pi_i^{-1}\left(\frac{\partial\mu_i}{\partial_\beta}\right). \end{aligned}$$

To find the estimates $\widehat{\beta}_n^c$, we use the Newton-Raphson algorithm. Denote $\widehat{\mu}_{in} = \{\widehat{\mu}_{i1n}, \widehat{\mu}_{i2n}, \ldots, \widehat{\mu}_{imn}\}^T$, where $\widehat{\mu}_i = \Phi(X_i\widehat{\beta}_n^c)$. Let $\widehat{\Pi}_{in}$ denote the estimator of Π_i obtained by substituting $\widehat{\mu}_{ijn}$ for μ_{ij}. We estimate $H(\beta)$ and $J(\beta)$ as

$$\begin{aligned} \widehat{H}_n &= n^{-1}\textstyle\sum_{i=1}^{n}(\frac{\partial\mu_i}{\partial_\beta}\Big|_{\widehat{\beta}_n^c})^T\widehat{\Pi}_{in}^{-1}(\frac{\partial\mu_i}{\partial_\beta}\Big|_{\widehat{\beta}_n^c}) \\ \hat{J}_n &= n^{-1}\textstyle\sum_{i=1}^{n}(\frac{\partial\mu_i}{\partial_\beta}\Big|_{\widehat{\beta}_n^c})^T\widehat{\Pi}_{in}^{-1}\ \widehat{\text{cov}}_n(y_i)\ \widehat{\Pi}_{in}^{-1}(\frac{\partial\mu_i}{\partial_\beta}\Big|_{\widehat{\beta}_n^c}), \end{aligned}$$

calculating the empirical variance as $\widehat{\text{cov}}_n(y_i) = (y_i - \widehat{\mu}_{in})(y_i - \widehat{\mu}_{in})^T$.

3.3 Quadratic exponential model

The quadratic exponential model is a popular tool used to model clustered binary data with intra-cluster interactions [6]. In this model, the binary observations take values $y_{ij} \in \{-1, 1\}$ and the joint distribution is given by

$$f_Y(y_i) \;\propto\; \exp\left\{\textstyle\sum_{j=1}^{m_i}\mu_{ij}^* y_{ij} + \sum_{j<j'} w_{ijj'}^* y_{ij}y_{ij'}\right\}, \qquad (7)$$

where μ^*_{ij} is a parameter which describes the main effect of the measurements and $w^*_{ijj'}$ describes the association between pairs of measurements within the cluster y_i. Independence corresponds to the case that $w^*_{ijj'} = 0$ and positive or negative correlation corresponds to $w^*_{ijj'} > 0$ or $w^*_{ijj'} < 0$, respectively. For simplicity, we consider the case that $\mu^*_{ij} = \mu^*_i$ and $w^*_{ijj'} = w^*_i$, noting that our methodology can be readily applied to the general scenario as well. Under this simplification, Molenberghs and Ryan [24], showed that the joint distribution can be equivalently written in terms of $z_i = \sum_{j=1}^{m_i} 1(y_{ij} = 1)$ (the number of successes in the i th cluster) as $f_Y(y_i) \propto \exp\{\mu_i z_i - w_i z_i(m_i - z_i)\}$, where $w_i = 2w^*_i$ and $\mu_i = 2\mu^*_i$.

Specifying the normalizing constant in eq. (7) is computationally difficult, but also necessary for the full likelihood evaluation. It is therefore desirable to use an alternative approach, one which does not involve such an intensive calculation. Replacing the joint distribution with the conditional distributions leads to a conditional composite likelihood function $cl(\mu, w; Y) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \log f(y_{ij}|\{y_{ij'}\}, j' \neq j)$, which does not require computation of the normalizing constant. We now define two conditional probabilities

$$p_{is} = \frac{\exp\{\mu_i - w_i(m_i - 2z_i + 1)\}}{1 + \exp\{\mu_i - w_i(m_i - 2z_i + 1)\}}, \qquad p_{if} = \frac{\exp\{-\mu_i + w_i(m_i - 2z_i - 1)\}}{1 + \exp\{-\mu_i + w_i(m_i - 2z_i - 1)\}}.$$

Heuristically, p_{is} is the conditional probability of one more success, given $z_i - 1$ successes and $m_i - z_i$ failures, while p_{if} is the conditional probability of one more failure, given z_i successes and $m_i - z_i - 1$ failures. Note that $p_{if} \neq 1 - p_{is}$, because of the term $m_i - 2z_i \pm 1$. The composite likelihood can now be expressed as $cl(\mu, w; Y) = \sum_{i=1}^{n} \left(z_i \log p_{is} + (m_i - z_i) \log p_{if}\right)$. This special form of the composite likelihood means that a logistic regression algorithm can be used to estimate the parameters. We model a covariate effect by using the linear model $\mu_i = X_i\beta$, with $w_i = w$ interpreted as an additional parameter. That is, for the parameter w, the value of the covariate is set to $-(m_i - 2z_i + 1)$ when $y_{ij} = 1$ and $-(m_i - 2z_i - 1)$ when $y_{ij} = -1$. This allows us to obtain CMLE estimates of both β and w using iterative re-weighted least squares, commonly used to solve logistic regression maximization problems.To estimate the covariance of $\widehat{\beta}^c_n$, we computed $\hat{J}_n$ as the empirical variance of the score vector, plugging in estimates of μ^*_i, w^* throughout. The Hessian matrix $\widehat{H}_n$ is estimated using the result from fitting the logistic model in R, see Geys et al. [6].

4 Simulation results

We evaluate the validity of our proposed approach on three different multivariate models from Section 3 using simulations. We randomly simulate covariates from independent normal distributions. We consider different values of the regression parameters and correlation structures. We simulate multivariate clustered data under varying cluster sizes and varying number of clusters. We perform multiple comparisons on the regression parameters using the proposed composite likelihood method. For the multivariate normal case, we also compare the proposed method with the likelihood approach.

We test two different global null hypotheses on the regression coefficients $\beta_1, \cdots, \beta_p$: many-to-one comparisons, $H_0 : \cap_{i=2}^{p}\{\beta_1 = \beta_i\}$; and all pairwise comparisons $H_0 : \cap_{1\leq i\neq j\leq p}\{\beta_i = \beta_j\}$. The results for many-to-one comparisons are summarized here while the results for all pairwise comparisons are provided in the supplementary material. We choose a collection of multiple testing methods including one-step Bonferroni method and the MNQ method based on the multivariate distribution of the test statistics. The equi-coordinate critical values for multivariate normal and multivariate t distributions are obtained using the R package mvtnorm [25]. We also examine the performance of Dunn-Sidák method, Holm's stage-wise procedure, and Scheffé's projection method and the results are provided in the supplementary material.

Part of our goal is to show practitioners what happens if the correlation structure in the clustered data is ignored. To this end, we also include a "misspecified" scenario, where independence is erroneously assumed within the clusters. Due to the specific composite likelihood methods we use (univariate marginals and univariate conditionals), such a misspecification is equivalent to $H(\theta) = J(\beta)$ in eq. (1). This results in an estimate of $\hat{\Gamma}_n = \widehat{H}_n^{-1}$. This misspecified scenario is included for comparison, and we consider it only with the MNQ multiple comparison method (that is, the MNQ cutoff is calculated based on V estimated by plugging in $\hat{\Gamma}_n = \widehat{H}_n^{-1}$). Throughout, it is referred to as the "naive" approach.

In our simulations, we study the three models described in the previous section. For each model, a different sample size is needed for our asymptotic approximations to be valid. We determine this sample size with an initial simulation, before we proceed with our more in-depth investigations. For each simulation setting, 10 000 simulated data sets are generated and the FWER is set to 0.05. The standard deviation for the observed FWER is hence approximately 0.002. These preliminary simulation results are given in Table 1. We observe that $n =$

200, 500 and 700 are required for the multivariate normal, multivariate probit and quadratic exponential models to maintain FWER within two standard deviations away from 0.05, respectively. These are the sample sizes used for the simulation results which follow.

Table 1 FWER under different sample sizes.

Model	**Sample size**					
	50	**200**	**500**	**700**	**1,000**	**4,000**
Multivariate normal	0.0544	0.0509	0.0492	0.0483	0.0495	0.0500
Multivariate probit	0.1140	0.0576	0.0501	0.0511	0.0506	0.0511
Quadratic exponential	0.1180	0.0580	0.0543	0.0519	0.0520	0.0504

To evaluate the power of each methods, we consider two different alternative scenarios: one alternative configuration a_1 with only one non-zero parameter having a large effect size, and a second alternative configuration a_2 with five true non-zero parameters but having small effect sizes for all. We are interested in the ability of the test to reject the global null hypothesis, but also in the ability of the test to reject the individual null hypotheses. Under the alternative scenario a_1, we calculate the power to reject the global hypothesis (denoted as "a_1" in the tables) and for the alternative configuration a_2, we calculate both the power to reject the global null hypothesis (denoted as "a_2" in the tables) and the average individual powers for the five individual true alternatives (denoted as "ind a_2" in the tables).

Table 2 Simulations results for three multivariate models.

					Normal			**Probit**			**Quad Exp**	
	ρ	m	p	**MNQ**	**naive**	**Bonf**	**MNQ**	**naive**	**Bonf**	**MNQ**	**naive**	**Bonf**
FWER	0	4	10	0.0545	0.0553	0.0419	0.0530	0.0506	0.0413	0.0514	0.0562	0.0400
a_1				0.8164	0.8166	0.7894	0.8700	0.8705	0.8477	0.5390	0.5534	0.5010
a_2				0.8057	0.8053	0.7617	0.9114	0.9109	0.8885	0.7067	0.7240	0.6573
ind a_2				0.1816	0.1816	0.1683	0.2166	0.2156	0.2039	0.1956	0.2057	0.1765
FWER			20	0.0511	0.0502	0.0352	0.0528	0.0503	0.0389	0.0561	0.0767	0.0404
a_1				0.7487	0.7476	0.7062	0.8258	0.8232	0.7902	0.4551	0.4853	0.3990
a_2				0.7150	0.7134	0.6687	0.8547	0.8511	0.8149	0.6040	0.6365	0.5237
ind a_2				0.1540	0.1535	0.1416	0.1887	0.1878	0.1769	0.1546	0.1669	0.1303
FWER		10	10	0.0479	0.0471	0.0375	0.0526	0.0515	0.0423	0.0491	0.0549	0.0381
a_1				0.9983	0.9983	0.9979	0.9996	0.9996	0.9995	0.5391	0.5535	0.5010
a_2				0.9993	0.9993	0.9986	1.0000	1.0000	1.0000	0.7066	0.7239	0.6573
ind a_2				0.2964	0.2963	0.2844	0.3330	0.3319	0.3168	0.1956	0.2057	0.1765
FWER			20	0.0487	0.0485	0.0363	0.0527	0.0508	0.0364	0.0561	0.0767	0.0404
a_1				0.9981	0.9980	0.9967	0.9993	0.9995	0.9985	0.4548	0.4849	0.3989
a_2				0.9978	0.9977	0.9963	1.0000	0.9999	0.9997	0.5971	0.6309	0.5255
ind a_2				0.2688	0.2681	0.2552	0.2973	0.2956	0.2811	0.1536	0.1663	0.1305
FWER	0.5	4	10	0.0513	0.0977	0.0390	0.0508	0.0793	0.0393	0.0521	0.0000	0.0417
a_1				0.7235	0.8129	0.6867	0.8102	0.8601	0.7808	0.7864	0.0307	0.7546
a_2				0.6922	0.8042	0.6615	0.8530	0.9038	0.8305	0.9050	0.0444	0.8800
ind a_2				0.1514	0.1878	0.1442	0.1970	0.2206	0.1864	0.3027	0.0090	0.2820
FWER			20	0.0510	0.1029	0.0377	0.0585	0.0915	0.0406	0.0509	0.0000	0.0377
a_1				0.6420	0.7526	0.5950	0.7578	0.8196	0.7082	0.7214	0.0158	0.6739
a_2				0.6031	0.7322	0.5437	0.7891	0.8534	0.7365	0.8460	0.0148	0.7998
ind a_2				0.1274	0.1607	0.1135	0.1727	0.1942	0.1571	0.2580	0.0030	0.2306
FWER		10	10	0.0520	0.2079	0.0410	0.0513	0.1437	0.0402	0.0521	0.0000	0.0417
a_1				0.9570	0.9936	0.9466	0.9900	0.9979	0.9871	0.7864	0.0307	0.7546
a_2				0.9555	0.9982	0.9438	0.9952	0.9997	0.9966	0.9141	0.0407	0.8800
ind a_2				0.2383	0.3381	0.2273	0.2815	0.3585	0.2704	0.3065	0.0083	0.2820
FWER			20	0.0459	0.2271	0.0328	0.0543	0.1622	0.0382	0.0509	0.0000	0.0378
a_1				0.9403	0.9938	0.9224	0.9862	0.9974	0.9784	0.7202	0.0161	0.6731
a_2				0.9222	0.9948	0.8932	0.9935	0.9998	0.9894	0.8460	0.0148	0.7998
ind a_2				0.2118	0.3072	0.1981	0.2575	0.3250	0.2450	0.2580	0.0030	0.2306

Table 3 Comparison of CMLE and MLE for multivariate normal model with exchangeable Σ.

	Method	m	p	ρ					
					0.2			**0.5**	
				MNQ	**naive**	**Bonf**	**MNQ**	**naive**	**Bonf**
FWER	CMLE	4	10	0.0494	0.0670	0.0389	0.0513	0.0977	0.0390
a_1				0.7760	0.8113	0.7453	0.7235	0.8129	0.6867
FWER	MLE			0.0492	0.0493	0.0443	0.0551	0.0110	0.0431
a_1				0.8319	0.8200	0.8053	1.0000	0.8423	0.9309
FWER	CMLE		20	0.0533	0.0734	0.0390	0.0510	0.1029	0.0377
a_1				0.7044	0.7490	0.6591	0.6420	0.7526	0.5950
FWER	MLE			0.0481	0.0475	0.0393	0.0591	0.0109	0.0445
a_1				0.7867	0.7581	0.7344	1.0000	0.7743	0.9070
FWER	CMLE	10	10	0.0467	0.1019	0.0357	0.0520	0.2079	0.0410
a_1				0.9912	0.9974	0.9875	0.9570	0.9936	0.9466
FWER	MLE			0.0536	0.0381	0.0364	0.0578	0.0071	0.0479
a_1				0.9994	0.9991	0.9992	1.0000	0.9998	1.0000
FWER	CMLE		20	0.0468	0.1057	0.0320	0.0459	0.2271	0.0328
a_1				0.9868	0.9970	0.9813	0.9403	0.9938	0.9224
FWER	MLE			0.0490	0.0388	0.0347	0.0674	0.0060	0.0529
a_1				0.9991	0.9987	0.9986	1.0000	0.9997	1.0000

Table 4 Simulations results with small sample sizes.

			Normal			**Probit**			**Quad Exp**	
	n	**MNQ**	**naive**	**Bonf**	**MNQ**	**naive**	**Bonf**	**MNQ**	**naive**	**Bonf**
FWER	50	0.0531	0.0962	0.0419	0.0839	0.0837	0.0674	0.1139	0.0003	0.0912
a_1		0.1750	0.2645	0.1502	0.1353	0.1411	0.1139	0.1551	0.0015	0.1273
a_2		0.1760	0.2628	0.1496	0.2849	0.3034	0.2398	0.1810	0.0034	0.1480
FWER	100	0.0474	0.0931	0.0379	0.0631	0.0808	0.0496	0.0704	0.0000	0.0551
a_1		0.3751	0.4836	0.3382	0.1832	0.2203	0.1594	0.1458	0.0004	0.1190
a_2		0.3563	0.4754	0.3181	0.5159	0.5835	0.4604	0.2016	0.0007	0.1703

Table 5 Simulations results using multivariate t approximation with $n = 50$.

Model		**MNQ**	**naive**	**Bonf**
Normal	FWER	0.0509	0.0896	0.0509
	a_1	0.1513	0.2245	0.1547
	a_2	0.1470	0.2241	0.1503
Probit	FWER	0.0667	0.0660	0.0673
	a_1	0.1112	0.1130	0.1152
	a_2	0.1206	0.1218	0.1143
	$ind\, a_2$	0.0227	0.0237	0.0214
Quad exp	FWER	0.0934	0.0001	0.0912
	a_1	0.1305	0.0008	0.1269
	a_2	0.1629	0.0013	0.1600

4.1 Multivariate normal model

We consider the multivariate normal model with $n = 200$ clusters, cluster size $m = 4$ or 10, and the number of covariates set to $p = 10$ or 20. Different Σ scenarios are considered: 1) exchangeable structures with $\sigma^2 = 0.8$ and $\rho = \mathrm{cov}(y_{ij}, y_{ik}) = 0$, or 0.5; 2) one arbitrary structure, where $\Sigma = ((1.3, 0.9, 0.5, 0.3)^T, (0.9, 1.9, 1.3, 0.3)^T, (0.5, 1.3, 1.3, 0.1)^T, (0.3, 0.9, 0.1, 0.7)^T)$. In each simulation, the $m \times p$ covariate matrix X_i is obtained by randomly sampling from normal distributions.

We consider here many-to-one comparisons where the first parameter is taken as the baseline. Under the global null hypothesis H_0, the true value of the regression parameters is set to $\beta^T = 0$, and the power is calculated under two different alternative configurations $\beta_{a_1}^T = (0, 0, 0, 0.032, 0, \dots, 0)$ and $\beta_{a_2}^T = (0, 0.008, 0.01, -0.03, 0.005, -0.01, 0, \dots, 0)$. Under β_{a_1}, there is only one true alternative, and we evaluate the power to reject the global null hypothesis. Under β_{a_2}, there are five true alternatives and we evaluate both the power to reject the global null and the average of five powers to reject the five true alternatives.

Table 2 summarizes the results of our simulations. Overall, it is shown that the method which utilizes MNQ and correctly accounts for the intra-cluster correlations, has the best performance. A comparison of MNQ and naive MNQ clearly shows the cost of ignoring these correlations: the FWER of MNQ is superior to that of naive MNQ for $\rho \neq 0$ (when $\rho = 0$ the two methods are almost identical). Notably, the power of the naive MNQ is occasionally higher than that of MNQ, however, this is only due to the over-inflation of the naive MNQ's FWER. Overall, MNQ exhibits the best performance among all of the multiple comparison procedures.

We also evaluate the efficiency of the maximum composite likelihood estimator versus the maximum likelihood estimator. In Table 3, it is observed that both type of statistics have very similar performance maintaining the type I error rates, while the method based on the composite likelihood estimator suffers some loss of power. For example, when $m = 10$, $p = 20$, $\rho = 0.5$, the power of MNQ based on the composite likelihood estimator is 0.94 compared to the power of 1.00 based on the maximum likelihood estimator. The MLE is obtained by treating the var-covariance matrix as a nuisance parameter. With some initial estimate for Σ, we maximize the loglikelihood for β. Then we estimate Σ using the residual covariance matrix. We iterate between the two steps until the estimates for β converges. For the maximum likelihood estimate, we use the matrix $n^{-1}(\sum_{i=1}^{n} X_i^T \hat{\Sigma} X_i)$ as the estimated covariance matrix for $\hat{\beta}$. In contrast, the naive method on the maximum likelihood estimate uses the matrix $n^{-1}(\sum_{i=1}^{n} X_i^T X_i)$ as the incorrect covariance matrix for $\hat{\beta}$.

It is observed that with increasing ρ and p for the multivariate distribution of the clustered data, the power of Bonferroni was not substantially smaller. The increase of ρ will increase the variability of each estimate $C^{(l)T}\hat{\beta}$ and hence decrease the power. When ρ increases from 0 to 0.5, we observe about 10% increase in the variability of the estimates and this is in compatible with the 5-10% power loss that we observe. We also conduct simulations with smaller sample sizes $n = 50$, and $n = 100$. It is shown that with n greater than 50, the statistics based on the plug-in estimate of the Godambe information matrix has satisfactory performance. Table 4 shows for $n = 50$, MNQ and Bonferroni maintains the FWER only for normal distribution, whereas for other two distributions, MNQ and Bonferroni tend to be liberal. The control of FWER is greatly improved with $n = 100$ for all three multivariate distributions. As the multivariate distribution of T_n can be approximated as a multivariate t distribution with n degrees of freedom, we conduct simulations to investigate the multivariate t approximation. Table 5 shows that the multivariate t approximation provides improved control of the FWER for normal, probit and quadratic exponential distribution compared to multivariate normal approximation (Table 4) with the same sample of $n = 50$.

4.2 Multivariate probit model

Here, we consider $n = 500$ clusters with a cluster size $m = 4$, or 10. The binary variables are generated by dichotomizing latent multivariate normal variables with a threshold of zero. For each cluster, an $m \times p$ covariate matrix X_i, with $p = 10$ or 20, is obtained by randomly sampling from normal distributions. The regression coefficient vector under the global null hypothesis is set to $\beta^T = 0$ and the two alternative configurations are $\beta_{a1}^T = (0, 0, 0, 0.03, 0, \dots, 0)$ and $\beta_{a2}^T = (0, 0.008, 0.01, -0.03, 0.005, -0.01, 0, \dots, 0)$. The latent multivariate random vector has a mean $X_i\beta$ and a correlation matrix with ρ on the off-diagonals and $\sigma = 1$. Here, we consider $\rho = 0$, or 0.5. The empirical results are given in Table 2. Similarly to the multivariate normal setting, the naive MNQ for the multivariate probit model has large FWER when $\rho = 0.5$. The MNQ method has better performance than the Bonferroni and naive method.

4.3 Quadratic exponential model

Here, we take a total of $n = 700$ clusters, and $p = 10$ or 20 predictors. The number of observations within each clusters, m_i, varies between clusters and is uniformly sampled from $\{4, 5, 6, 7, 8\}$. The $m_i \times p$ covariate matrix X_i is sampled from a standard normal distribution. We also consider two different values for the interaction parameter: $w = 0$ or 0.5. The null value of the regression coefficients is $\beta^T = 0$ and the two alternative configurations are to $\beta_{a1}^T = (0, 0, 0, 0.12, 0, \dots, 0)$ and $\beta_{a2}^T = (0, 0.08, 0.12, -0.03, 0.05, -0.08, 0, \dots, 0)$. The empirical FWER and power are computed and summarized in Table 2. Overall, MNQ has clearly the best performance.

5 Analysis of depression data

Table 6 Composite likelihood estimates of the health factors' regression coefficients.

	Estimate	**SE**	**Unadjusted p-value**
Intercept	– 2.1751	0.0508	$< 2e-16$
Sleeplessness	1.3330	0.0290	$< 2e-16$
Smoking	0.2826	0.0439	$< 2e-16$
High blood pressure	0.0764	0.0219	$2.07e-11$
Diabetes	0.0710	0.0296	$8.96e-07$
Difficulty in walking	0.0695	0.0054	$< 2e-16$
Age	0.0093	0.00003	$< 2e-16$
Activity	– 0.0156	0.0064	$2.35e-05$
w	0.2877	0.0094	$< 2e-16$

We apply our proposed method to the health and retirement study (HRS) dataset. Information about health, financial situation, family structure, and health factors were collected by the RAND center at the University of Michigan. Depression status is recorded as a binary response variable. Seven health factors include "age" (in months), "smoking", "restless sleep", "diabetes", "high blood pressure", "activity", and "difficulty in walking" are considered as predictors and each predictor is highly significant with unadjusted p-values all less than 10^{-4}. The effect size estimates and unadjusted p-values are reported in Table 6. As we are interested in comparing the importance of different predictors, we perform multiple comparisons on the effects of these seven health factors. For each individual we include only the years for which all of the factors were recorded. In total, there are $33,636$ people included in the analysis and the number of repeated measurements vary across individuals. As the response variable is binary and the cluster sizes vary, we propose to use a quadratic exponential model to model this data set. The w parameter in the quadratic exponential model allows us to account for the interaction effect among the repeated measurements for the same individuals. The full likelihood approach is computationally challenging for this model, and hence we use the proposed composite likelihood method to perform inference.

To compare the effect sizes of all the seven health factors, we perform both all pairwise comparisons and many-to-one comparisons on the seven parameters. For all pairwise comparisons, the MNQ approach rejects 15 of the 21 hypothesis. The results are given in Table 7. Based on the estimates of the effect sizes (Table 6), we note that "restless sleep" and "smoking" are the two health factors with the largest effect sizes. The pairwise comparisons between restless sleep with all other health factors and smoking with all other factors are rejected. This shows that "restless sleep" and "smoking" are the two leading health factors for the occurrence of depression. "High blood pressure", "diabetes", and "difficulty in walking" have less influence on the occurrence rate of depression. When we examine the three pairwise comparisons among these three factors, the three null hypotheses are not rejected, indicating that these three health factors have similar effects on the disease. Furthermore, when we compare "high blood pressure" with "age" and "activity", the test rejects the two comparisons, indicating that "high blood pressure" is more important than "age" and "activity" with regard to the disease development.

To show how different the results will be if the within-patient correlations are ignored, we also compare the result of the MNQ with the naive MNQ method. Both the MNQ and the naive MNQ reject the global null hypothesis that all pairs of health factors have equal effects on the depression status. The MNQ method rejects 15 hypotheses, whereas the naive MNQ method rejects 18 out of the total 21 hypotheses. MNQ and naive MNQ are in agreement in all the aforementioned comparisons. However, when we compare the effect sizes between age and diabetes, diabetes and activity, age and activity, the MNQ method cannot reject these three null hypotheses while the naive method rejects all three. The difference between the two methods is due to the correlation among the repeated measurements, which is estimated as $\widehat{w} = 0.2877$. By ignoring this correlation, as in the naive method, the standard errors are underestimated leading to more rejections.

We also conduct many-to-one comparisons on the seven health factors using "diabetes" as the baseline factor. Table 8 shows that the comparisons are consistent with our findings in Table 7. "Restless sleep" and "smoking" are two factors which are much more influential than "diabetes" in terms of increasing the risk of depression, whereas "activity" is one factor which is more important than "diabetes" in terms of decreasing the risk of depression. The "age", "high blood pressure" and "difficulty in walking" seem to be have similar effect sizes as "diabetes". Naive MNQ method rejects one more comparison than MNQ method indicating that naive method is also more liberal in many-to-one comparisons.

Table 7 Results of MNQ and naive MNQ for all pairwise comparisons on the depression study data set. (A: fail to reject, R: reject H_0).

H_0	**MNQ**	**naive**	H_0	**MNQ**	**naive**
$\beta_{sleep} = \beta_{smoke}$	R	R	$\beta_{hbp} = \beta_{diabet}$	A	A
$\beta_{sleep} = \beta_{hbp}$	R	R	$\beta_{hbp} = \beta_{diff\ walk}$	A	A
$\beta_{sleep} = \beta_{diabet}$	R	R	$\beta_{hbp} = \beta_{age}$	R	R
$\beta_{sleep} = \beta_{diff\ walk}$	R	R	$\beta_{hbp} = \beta_{activity}$	R	R
$\beta_{sleep} = \beta_{age}$	R	R	$\beta_{diabet} = \beta_{diff\ walk}$	A	A
$\beta_{sleep} = \beta_{activity}$	R	R	$\beta_{diabet} = \beta_{age}$	A	R
$\beta_{smoke} = \beta_{hbp}$	R	R	$\beta_{diabet} = \beta_{activity}$	A	R
$\beta_{smoke} = \beta_{diabet}$	R	R	$\beta_{diff\ walk} = \beta_{age}$	R	R
$\beta_{smoke} = \beta_{diff\ walk}$	R	R	$\beta_{diff\ walk} = \beta_{activity}$	R	R
$\beta_{smoke} = \beta_{age}$	R	R	$\beta_{age} = \beta_{activity}$	A	R
$\beta_{smoke} = \beta_{activity}$	R	R			

Table 8 Results of MNQ and naive MNQ for many-to-one comparisons on the depression study data set. (A: fail to reject, R: reject H_0).

H_0	**MNQ**	**naive**
$\beta_{dibet} = \beta_{sleep}$	R	R
$\beta_{dibet} = \beta_{smoke}$	R	R
$\beta_{dibet} = \beta_{hbp}$	A	A
$\beta_{dibet} = \beta_{diff\ walk}$	A	A
$\beta_{dibet} = \beta_{age}$	A	R
$\beta_{dibet} = \beta_{activity}$	R	R

6 Discussion

For many correlated multivariate datasets, it is often computationally convenient to perform multiple comparisons based on the composite likelihood. Theory is developed based on the asymptotic properties of the composite likelihood test statistics. Sample sizes greater than 50 will be sufficient to apply the proposed multiple testing procedures in various models examined in the simulations. We establish the strong control of the FWER of our proposed composite likelihood test statistics with closed testing procedures. The method is illustrated for three different models: multivariate normal, multivariate probit and quadratic exponential. The equi-coordinate quantile of a multivariate normal distribution is used as a threshold for test statistics compared to some well-known traditional methods. This MNQ method, which is based on composite likelihood test statistics and uses multivariate normal quantiles to derive cut-off values for the test statistics, shows a better control of the FWER in most simulation settings, compared to the other test procedures.

7 Supplementary Files

We provide a supplementary file containing additional simulation results and technical derivations. An R package named CLMC is developed and can be downloaded from github account “m-azad”.

Acknowledgement

This work is supported by NSERC grants held by Gao and Jankowski.

References

[1] Barua A, Ghosh MK, Kar N, Basiliod MA.. Prevalence of depressive disorders in the elderly, 2011.
[2] Cox DR, Reid N.. A note on pseudolikelihood constructed from marginal densities. Biometrika. 2004;91:729–737.
[3] Lindsay BG.. Composite likelihood methods. Contemporary Mathematics. 1988;80:221–239.
[4] Varin C.. On composite marginal likelihoods. AStA Adv Stat Anal. 2008;92:1–28.
[5] Varin C, Reid N, Firth D.. An overview of composite likelihood methods. Stat Sin. 2011;21:5–42.
[6] Geys H, Molenberghs G, Ryan LM.. Pseudo-likelihood inference for clustered binary data. Commun Stat Theory Methods. 1997;26:2743–2767.
[7] Renard D, Molenberghs G, Geys H.. A pairwise likelihood approach to estimation in multilevel probit models. Comput Stat Data Anal. 2004;44:649–667.
[8] Zhao Y, Joe H.. Composite likelihood estimation in multivariate data analysis. Canadian J Stat. 2005;33:335–356.
[9] Bretz F, Hothorn T, Westfall P.. Multiple comparisons using R FL:Chapman and Hall/CRC Press, 2010 Boca Raton.
[10] Hochberg Y, Tamhane A.. Multiple comparison procedures. New York: Willy, 1987.
[11] Sidak Z.. On multivariate normal probabilities of rectangles: Their dependence on correlations. Ann Math Stat. 1968;39:1425–1434.
[12] Schéffe. The analysis of variance. Wiley, 1959 New York.
[13] Simes RJ.. An improved Bonferroni procedure for multiple tests of significance. Biometrika. 1986;73:751–754.
[14] Holm S.. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6:65–70.
[15] Hommel G.. A stagewise rejective multiple test procedure based on a modified bonferroni test. Biometrika. 1988;75:383–386.
[16] Hothorn T, Bretz F, Westfall P.. Simultaneous inference in general parametric models. Biom J. 2008;50:346–363.
[17] Konietschke F, Bosiger S, Brunner E, Hothorn LA.. Are multiple contrast tests superior to the anova?. Int J Biostat. 2013;9:11.
[18] Konietschke F, Hothorn LA, Brunner E.. Rank-based multiple test procedures and simultaneous confidence intervals. Electron J Stat. 2012;6:738–759.
[19] Besag J.. Spatial interaction and the statistical analysis of lattice systems. 1974;36:192–236 J Roy Stat Soc Ser BWith discussion byand with a reply by the author.
[20] Xu X, Reid N.. On the robustness of maximum composite likelihood estimate. J Stat Plann Inference. 2011;141:3047–3054.
[21] Varin C, Vidoni P.. A note on composite likelihood inference and model selection. Biometrika. 2005;92:519–528.
[22] Gabriel KR.. Simultaneous test procedures, some theory of multiple comparisons. Ann Math Stat. 1969;40:224–250.
[23] Marcus R, Peritz E, GK R. On closed testing procedures with specific reference to ordered analysis of variance. Biometrika. 1976;63:655–660.
[24] Molenberghs G, Ryan LM.. An exponential family model for clustered multivariate binary data. Environmetrics. 1999;10:279–300.
[25] Hothorn T, Bretz F, Westfall P, Heiberger RM, Schutzenmeister A.. multcomp: Simultaneous inference for general linear hypotheses. *R package version*. 1.1-7,, 2010. Available at http://CRAN.R-project.org/package=multcomp.

Supplemental Material: The online version of this article (DOI:10.1515/ijb-2016-0004) offers supplementary material, available to authorized users.