ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

Working Correlation Structures: Application in Liver Cirrhosis Therapeutic Effect Trial

Çalışma Korelasyon Yapıları: Karaciğer Sirozunda Uygulama Tedavisel Etki Denemesi

Ramesh K. VISHWAKARMA,[a]
Dilip C. NATH,[a]
Atanu BHATTACHARJEE[b]

[a]Department of Statistics, Gauhati University, Guwahati
[b]Division of Clinical Research and Biostatistics, Malabar Cancer Centre, Thalassery

Geliş Tarihi/*Received:* 27.02.2016
Kabul Tarihi/*Accepted:* 12.04.2016

Yazışma Adresi/*Correspondence:*
Atanu BHATTACHARJEE
Malabar Cancer Centre,
Division of Clinical Research and Biostatistics,
Thalassery, INDIA
atanustat@gmail.com

ABSTRACT Objective: Correlated responses are common in repeatedly measured clinical trial data. The generalized estimating equations (GEE) method is popular for analyzing correlated responses. It is important to select a proper working correlation matrix because an inappropriate choice will lead to inefficient parameter estimation. In this paper, we examine criterion of quasi-likelihood information criterion (QIC) for selecting a working correlation structure, and have compared with the performance of the correlation information criteria (CIC) of the correlation structures on liver cirrhosis patients. **Material and Methods:** The computation code for CIC is performed into open source software R. The covariates like therapy and visit are used to predict Meld scores (It is continuous disease severities scale with highly predictive of the risk of dying from liver cirrhosis) in GEE model to examine the performance of QIC and CIC after considering three different working correlation structures. **Results:** The GEE model has been performed to compare QIC and CIC after considering three working correlation structure. In case of AR (1) correlation structure, it is found that similar regression parameter estimates are observed for both information criteria techniques. **Conclusion:** The study indicates that the CIC is useful for selecting appropriate correlation structures for liver cirrhosis data from phase III clinical trial.

Key Words: Linear models; statistics as topic; survival analysis; Bayes theorem; survival rate

ÖZET Amaç: İlişkili yanıtlar tekrar tekrar ölçülen klinik deneme verilerinde yaygındır. Genelleştirilmiş hesaplama denklemler (GEE) yöntemi ilişkili yanıtları analiz etmek için yaygındır. Uygunsuz bir seçim etkisiz parametre tahminine yol açacağından doğru bir çalışma korelasyon matrisi seçmek önemlidir. Bu makalede bir çalışma korelasyon yapısı seçmek için yarı-en çok olabilirlik bilgi kriterini (QIC) inceledik ve bu kriteri karaciğer sirozu hastalarındaki korelasyon yapılarının korelasyon bilgi kriterinin (CIC) performansı ile karşılaştırdık. **Gereç ve Yöntemler:** CIC için hesaplama kodu açık kaynak kodlu yazılım R'a uygulanmıştır. GEE modelinde üç farklı çalışma korelasyon yapıları değerlendirildikten sonra QIC ve CIC performansını incelemek için terapi ve ziyaret gibi ortak değişkenler Meld skorlarını (karaciğer sirozundan ölme riskinin yüksek derecede kestirimi ile sürekli hastalık şiddet dereceleri ölçeği) tahmin etmek için kullanılmaktadır. **Bulgular:** GEE modeli üç çalışma korelasyon yapısı incelendikten sonra QIC ve CIC'nin karşılaştırılması için uygulanmıştır. AR (1) korelasyon yapısı durumunda, her iki bilgi kriteri yöntemi için de benzer regresyon parametre tahminlerinin gözlendiği bulunmuştur. **Sonuç:** Çalışma evre III klinik deneme karaciğer siroz verileri için uygun korelasyon yapılarının seçilmesinde CIC'ın kullanışlı olduğunu göstermektedir.

Anahtar Kelimeler: Doğrusal modeller; konu olarak istatistikler; Bayes teoremi; sağkalım hızı

doi: 10.5336/biostatic.2016-51090

Turkiye Klinikleri J Biostat 2016;8(2):116-24

Liver Cirrhosis occurs due to scarring of the liver for abnormal liver function as a consequence of chronic (long-term) liver injury. It influences the gradual shrinkage of the size of the liver. In liver cirrhosis problem, the duration between transplantation to recover is the crucial period for patients.[1,2] The performance of patient's Liver is generally measured through follow-up periods with liver functioning effects. An excess amount of alcohol consumption and chronic infection with hepatitis B and hepatitis C virus are the common risk factor for liver Cirrhosis 1. However, there are several unknown factors for liver cirrhosis. In this study, the disease severity in the patient with liver cirrhosis is measured through the model for the end stage liver disease (MELD).[2] It is continuous disease severities scale with highly predictive of the risk of dying from the liver.The MELD score is considered as the standard of reference for the diagnosis and staging of liver Cirrhosis.[2] The different biochemical parameters viz. serum creatinine, bilirubin & INR (International Normalized Ratio) are used to calculate the MELD score. The details to calculate the MELD score can be cited with http://en.wikipedia.org/wiki/Model_for_End-Stage_Liver_Disease. The low MELD score is positively associated with mortality.[2,3] Correlation among repeated measurement of MELD scores indicates that independence can no longer be assumed. Therefore, most standard statistical analyses cannot be used to analyze this type of data.[3] If standard analyses (for example ANOVA test used without accounting for dependence within the subjects) are used, the likelihood of Type I errors will be increased. A number of approaches like repeated measure ANOVA, repeated measures ANCOVA etc. are available for analyzing correlated data. However, selecting which approach is the best to analyze a particular study is unimportant, because each of these methods has a different theoretical paradigm, and its own strengths and weaknesses.[4] The challenge is to specify the particular correlation structure to model the repeatedly measured data.[5] However, it is difficult to specify the correct correlation structure. Different types of correlation can be specified and checked through model selection criteria. The generalized estimating equation (GEE) is widely applied a tool to deal with correlated repeated data. The Akaike information criterion (AIC) is another widely used tool for model selection criteria, but it is failing to select the model of dependent MELD scores.[6] The extension of AIC is also available to deal with dependent observations by quasi-log-likelihood under the independence model information criteria (QIC).[7] Longitudinal data can be handled by marginal modeling and the choice is to apply trough GEE.[8] The joint estimates in the marginal models have been applied to children's heart function data through modified GEE.[9]

OBJECTIVE

The aim of this work is to examine the quasi-likelihood information criterion (QIC) for selecting working correlation structures and also to compare the performance of the correlation information criteria (CIC) of the working correlation structures by assessing the therapeutic effect of severity of liver Cirrhosis measured through MELD scores collected on follow-up visits during course of treatment. The response of treatment is considered as MELD score and therapy, visits as covariates of interest.

DATA METHODOLOGY

The main step for analysis of correlated data is to select the appropriate covariance structure, which explains the form (or structure) of the correlation data among time points within subjects. This is vital because the overall model fit, the parameter estimates, and their standard errors can

be sensitive to the model covariance structure. The covariance is given a simplifying structure, as this reduces the number of parameters and can improve model convergence.[4]

We define variance-covariance of the responses for MELD scores of ith subject $\mathrm{Var}(Y_i)$ using the m × m symmetric matrix. The diagonal elements of variances and the off-diagonal elements are covariances. There are a large number of covariance structures to choose from. In this paper, we focus on three working correlation structures: independent, exchangeable and autoregressive. These three structures cover a range of different scenarios for the pattern of covariance and are most commonly available in statistics packages like R, SAS SPSS. For example, we might assume that the covariance between all MELD scores from the same subject is constant and that the variance remains constant over time is known as the exchangeable covariance matrix because the MELD scores from any subject could be re-arranged (exchanged) in time, and the covariance between MELD scores would remain the same.

A number of different covariance structures are available that cover a range of assumptions about the associations between responses from the same subject. An independent covariance would be appropriate when none of the responses are correlated.[4] This is equivalent to the exchangeable covariance with covariances= 0. This structure is useful for determining whether more complex structures improves the model fit. An exchangeable covariance would be appropriate when responses from the same subject are equally correlated, regardless of the distance between responses. An autoregressive covariance would be appropriate when the correlation between responses decays with distance. It assumes a steady decay in correlation with increasing time or distance between MELD scores. It is common to use an autoregressive model of order one, labeled AR(1), which has one correlation parameter and one variance (as does the exchangeable covariance). The empirical data is considered from the path http://www4.stat.ncsu.edu/~boos/var.select/pbc.html; accessed on January 6, 2013. More details about data can be found with Nath et al.[2] A total of 175 Patients with cirrhosis are randomized in two different treatment group i.e., 86 in each group.

In the treated group, patients who consented for participation had received infusion treatment of human fetal liver progenitor cell (HFPLC) as well conventional medical treatment in the background. In a control group, patients who consented for participation had only received the conventional medical treatment.[2,3] The duration of the study was 36 months with 7 visits (Baseline, 3 months, 6 months, 9 months, 12 months, 24 months & 36 months) . The parameters under consideration were MELD Scores, therapy, and visit. Each patient MELD score are taken at seven different time points.

STATISTICAL ANALYSIS

We have used QIC, CIC information criteria for statistical analysis of liver cirrhosis patient data. The equations and terms used in QIC, CIC, and AIC for fitting GEE model are elaborated in Appendix I. We compare the performance of the two information criteria using data from liver cirrhosis phase III clinical trial (with known covariance structure). During the therapy of liver cirrhosis, each patient's MELD scores are observed till 7th visits. The explanatory variables namely drug therapy and visit are considered for GEE model. The drug therapy is captured as binary format into two levels. And visit is consist of seven categories namely "Baseline","Month 1", "Month 3", "Month 6", "Month 12", "Month 24" and "Month 36". The explanatory variables are denoted as X'= [xij1;xij 2,……., xij4]'.

APPENDIX I

1. Generalized Estimating Equation (GEE)

We start with few notations and assumptions. We label the repeated data from subject i using hence there are n_i responses for i^{th} individuals and the covariates is denoted as $X_i = (X_{i1}, X_{i2}, \ldots, X_{in_i})'$ (A1)

For simplicity we only consider Normally distributed response data (i.e., Y has a multivariate Normal distribution), and balanced data so each subject has the same number of responses n. We assume that the repeated data were measured on same subject at multiple times ($t = 1, \ldots, n$). However, the methods could be applied to non-longitudinal data, such as responses from the same samples that are spatially clustered.

Here, Y_{ij} are the scalars and X_{ij}' are corresponding vectors. Itis to be noted that Y_{i1} and Y_{i2} and dependent but Y_{ij} and Y_{Ij} for any value of j are independent. The simple linear regression is used to define the model as $g(\mu_i) = x_i\beta$, g is a link function and $\mu_i = E(Y_i|X_i)$ and the regression coefficients are $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$. The GEE is useful to estimates the parameters β.

Now, it can be defined as

$S(\beta, R, D) = \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - \mu_i) = 0$ (A2)

$D_i = D_i(\beta) = \delta\mu_i(\beta)/\delta\beta'$ (A3)

V_i is the working matrix of Y_i and

$V_i = A_i^{1/2} R(\alpha) A_i^{1/2}$ (A4)

The term A_i is the diagonal matrix with the elements $Var(Y_{ij}) = \varphi V(\mu_{ij})$, which is specified as a function of the mean μ_{ij}. α is the unknown parameters with the working correlation structures; it can be estimated through moments methods or with other estimating equations.

The advantage to use GEE is that it gives a consistent estimation about β aus $\hat{\beta}$. It can also perform consistent, even are (i.e., correlation structure) is miss-specified [5, 10]. In contrast to that the efficacy of wrong correlation specification is observed and found it affected by 40% due to wrong specification of correlation structure[11, 17]. So, the specification of proper correlation is important. The "Quasi-likelihood" based approach, other types of model selection criteria is specified below.

Quasi-likelihood

The AIC can only be used in association with mixed models, and this cannot be used with GEEs to choose either the best set of explanatory variables or covariance matrix, since GEE estimation is based on the quasi-likelihood rather than the maximum likelihood. The quasi-likelihood counterpart to the AIC is the QIC, or the "quasi-likelihood under the independence model information criterion" [4]. The QIC was derived from the AIC and is conceptually similar.

Let the response of interest is Y. Let the mean of the response is denoted as $\mu = E(y)$, and the regression parameter is β.

Further, $E(y) = \mu$ and $Var(y) = \varphi V(\mu)$. Here, the φ is the dispersion parameter. The function of quasi-likelihood is defined as

$Q(\mu, \varphi, y) = \int_y^{\mu} \frac{y - t}{\varphi V(t)} dt$ (A5)

The binary data are $y \sim Bin(n, \pi)$ and further

$V(\mu) = \mu(1 - \mu/n)$ (A6)

and $Q(\mu, \varphi, y) = L(\mu, \varphi; y)/\varphi$ (A7)

Here, $L(\mu, \varphi, y) = y\log[\mu/(n - \mu)] + n\log(n - \mu)$ (A8)

is the log-likelihood for a binomial distribution.[12] Let $\varphi = 1$, then the quasi-likelihood will reduce to L. Now in case of $\varphi > 1$, the problem can be handled by over dispersion.

Let the $1 \times p$ is a covariate x and it specified with regression model $E(y) = \mu = g^{-1}(x\beta)$ and $Var(y) = \varphi V(\mu)$. Further, the regression coefficients (β) is defined from the function

$Q(\beta, \varphi; (y, x)) = Q(g(x\beta)^{-1}, \varphi; y)$ (A9)

The prior defined term D is called as

$Q(\beta, \varphi; I, D) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} Q(\beta, \varphi; (Y_{ij}, X_{ij}))$ (A10)

The GEE above is defined as Quasilikelihood by the proving

$\frac{\delta Q(\beta, \varphi; I, D)}{\delta\beta} = S(\beta, R, D)$ (A11)

In this work the coefficients of regression parameters are estimated as $\hat{\beta}_{(R_i)}$, where the correlation structure is assumed as $R_i = R_i(\alpha)$. Now the independent correlation structure is defined as $R_i = I$ and estimated value obtained is denoted as $\hat{\beta}_{(I)}$. The estimated dispersion parameter obtained through GEE is $\hat{\varphi}(I)$ for dispersion raised due to assumption of independent correlation structure. The performance of QIC in different correlation structure has been explored [16, 18]. The performance of CIC is also being performed in the same scenario [2, 5]. However, CIC is not the alternative of QIC. But both can play a joint role for model computation.

Correlation Information Criterion(CIC)

Hin and Wang (2009) proposed CIC as a modification of QIC to improve its performance: CIC is constructed using by the second term which represents the penalty of QIC. The first term in QIC denotes the sum of quasi-likelihood for all MELD scores under the assumption that the subjects and time points are independent. It makes sense to ignore the first term when comparing different

working correlation structures, since the term mostly does not depend on the specified working correlation structure.

QIC is separated into T_1 and T_2 by

$$QIC = T_1 + T_2 \quad \text{(A12)}$$

Further, $T_1 = -2Q(\hat{\beta}, \varphi, T, D)$ and $T_2 = 2tr[\hat{\Omega}\hat{V}_r](\beta, \varphi, T, D)$ (A13)

Here, the sum of quasi-likelihood for the $\sum_{i=1}^{m} m_i$ observation in the data set D is denoted by T_1.

Now, $E(T_1)$ is free from R (correlation structure). But, the term $E(T_2)$ is not free from R. Now, the term T_1 can be computed by ignoring the correlation structure. The estimation of QIC is based on Kullback-Leibler method. Now, random error raised due to miss specification by random-error of T_2 for QIC. Here, T_1 assumed no correlation structure but T_2 assumed the correlation structure. The simplest modification of QIC is defined as

$$CIC = \text{tr}[(\sum_{i=1}^{n} D_i' A_i^{-1} D_i)\hat{V}_r|_{\beta=\hat{\beta}(R_i)}] \quad \text{(A14)}$$

Now, the V_r in the above equation is defined as

$$D_i' V_i^{-1} D_i' V_i^{-1} Var(y_i) V_i^{-1} D_i (D_i' V_i^{-1} D)_i^{-1}|_{\beta=\hat{\beta}(R_i), \phi=\hat{\phi}(R_i)} \quad \text{(A15)}$$

$$CIC = \text{tr}[\sum_{i=1}^{n} (D_i' A_i^{-1} D_i)^{-1} (D_i' V_i^{-1} D_i' V_i^{-1} \text{Var}(y_i) V_i^{-1} D_i (D_i' V_i^{-1} D)|_{\beta=\hat{\beta}(R_i), \phi=\hat{\phi}(R_i)}] \quad \text{(A16)}$$

And the corresponding expression of CIC becomes

$$CIC = \text{tr}[\sum_{i=1}^{n} (D_i' A_i^{-1} D_i)^{-1} [\sum_{i=1}^{n} (D_i' A_i^{-1} D_i)^{-1} (D_i' V_i^{-1} D_i' V_i^{-1} \text{Var}(y_i) V_i^{-1} D_i (D_i' V_i^{-1} D)|_{\beta=\hat{\beta}(R_i), \phi=\hat{\phi}(R_i)}] \lim_{n\to\infty} CIC = tr(\Omega V_R) \quad \text{(A17)}$$

when correlation structure is correctly specified then the right hand side of the above equation is formed as,

$$tr[\sum_{i=1}^{n} (D_i' A_i^{-1} D_i)(D_i' Var^{-1}(y_i) D_i)^{-1}] = \varepsilon + \text{tr}[\sum_{i=1}^{n} \{D_i' (A_i^{-1} - V_i^{-1}) D_i\} (D_i' V_i^{-1} D_i)^{-1}] \quad \text{(A18)}$$

Akaike information criterion (AIC)

Discussion about AIC can be found from [9]. It is defined as

$$AIC = 2k - 2ln(L) \quad \text{(A19)}$$

where L is the number maximized value of likelihood and k is the number of parameters in the model. The AIC value for all models can be calculated and minimum value obtained from the specific model is defined as best fitted model. AIC castigates models with larger numbers of parameters. Particularly, Vaida and Blanchard (2005) [10] extended the AIC for repeated measures data as conditional AIC by

$$AIC_1 = -2E(y, b) E_{y_0|b}[l\{y_0|\hat{\beta}(y), \hat{b}(y)\}] \quad \text{(A20)}$$

The function $l(.|.)$ is the conditional log-likelihood and $\hat{\beta}(y), and \hat{b}(y)$ are the estimated parameters. The unbiased part can be denoted as

$$AIC_1 = -2l\{y_0|\hat{\beta}(y), \hat{b}(y)\} + 2\rho \quad \text{(A21)}$$

The therapeutic effect is classified as:

$$X_{ij} = \begin{matrix} 1(\text{if therapy} = \text{HFPLC}) \\ 0(\text{if therapy} = \text{CONTROL}) \end{matrix} \quad (1)$$

Here, Yij is the MELD value of the ith patient in jth visit, j=1,…7.and μij=E(yij) is the mean of MELD value. The variance function of the binomial distribution is v(μij)= μij(1-μij), in the logit link function of $g(\mu_{ij}) = \log(\mu_{ij})/(1-\mu_{ij}))$. The model for the mean response is $g(\mu_{ij}) = x_{ij}'\beta$. The coefficient β is the vector of regression parameters. In the software R 3.0.3, the library "geepack is" used to specify the "working covariance" and variance function. The Generalized Estimating Equation (GEE) is applied to fit with link function in

$$\text{logit}[Y_{it}] = \beta_0 + \beta_1 \text{Therapy} + \beta_2 \text{Visit} + \beta_3 (\text{Therapy} * \text{Visit}) \quad (2)$$

The GEE is applied through specifying different types of correlation structures.

RESULT

Table 1 describes the demographic and baseline characteristics of 175 patients suffering from liver

TABLE 1: Baseline and demographic characteristic.

Parameters	Treatment group Mean (SD)	Control Group Mean (SD)
Age	48.60(9.38)	49.85(11.06)
Gender Male Female	10 (12.05%) 73 (87.95%)	70 (76.09%) 22 (23.91%)
Height	164.89(4.46)	165.38(5.84)
Weight	65.71(5.24)	69.51(8.78)
Respiratory rate (RR)	25.99(16.59)	21.43(1.62)
Heart Rate	72.89(15.61)	77.22(2.08)

TABLE 2: Estimates of different parameter under different correlation structure through QIC.

Correlation Structure	Variable (Parameter)	Estimation	S.E.	p-value
AR(1)	Intercept(β0)	0.72	0.15	0.11
	Therapy (β1)	0.23	0.08	0.12
	Visit (β2)	-0.40	0.09	0.92
	Therapy*Visit (β3)	0.22	0.04	0.03
Independent	Intercept (β0)	0.69	0.19	0.09
	Therapy (β1)	0.29	0.11	0.13
	Visit (β2)	-0.39	0.13	0.71
	Therapy*Visit (β3)	0.19	0.08	0.13
Exchangeable	Intercept (β0)	0.70	0.13	0.10
	Therapy (β1)	0.23	0.05	0.11
	Visit (β2)	-0.42	0.07	0.67
	Therapy*Visit (β3)	0.21	0.14	0.15

TABLE 3: Estimates of different parameter under different correlation structure through CIC.

Correlation Structure	Variable (Parameter)	Estimation	S.E.	p-value
AR(1)	Intercept(β0)	0.72	0.15	0.09
	Therapy (β1)	0.23	0.08	0.11
	Visit(β2)	-0.40	0.09	0.79
	Therapy*Visit (β3)	0.22	0.04	0.07
Independent	Intercept(β0)	0.69	0.19	0.10
	Therapy (β1)	0.29	0.11	0.14
	Visit(β2)	-0.39	0.13	0.65
	Therapy*Visit (β3)	0.19	0.08	0.15
Exchangeable	Intercept(β0)	0.70	0.13	0.12
	Therapy (β1)	0.23	0.05	0.14
	Visit(β2)	-0.42	0.07	0.68
	Therapy*Visit (β3)	0.21	0.14	0.17

cirrhosis. The mean age of liver cirrhosis patient in the treatment group is 48.6 with standard deviation (SD) 9.38 whereas in a control group, the mean age is 49.85 with SD 11.06. The distribution of male in treated and control group is 10 (12.05%) and 70 (76.09%) respectively. Similarly, the distribution of female in treated and control is 73 (87.95%) and 22 (23.91%) respectively. The mean height of liver cirrhosis patients in the treatment group is 164.89 with SD 4.46 whereas, in a control group, the mean height is 165.38 with SD 5.84. The mean weight of liver cirrhosis patient in the treatment group is 65.71 with SD 5.24 whereas, in a control group, the mean weight is 69.51 with SD 8.87. The mean Respiratory Rate (RR) of liver cirrhosis patient in the treatment group is 25.99 with SD 16.59 whereas in the control group, the mean RR is 21.43 with SD 1.62. The mean Hear Rate (HR) of liver cirrhosis patient in the treatment group is 72.89 with SD 15.61 whereas, in a control group, the mean HR is 77.22 with SD 2.08.

The parameter estimations under QIC criteria for all three working correlations structures are illustrated in Table 2. The GEE model was fitted for simultaneously with different correlation structures namely AR(1), independent, exchangeable using QIC approach for estimating the parameters. When we considered AR (1) correlation structure into the model, we found that the estimated difference in slopes, 0.22, is significant at 5% level of significance, indicating that the responses are increasing over time quickly for the therapy group.the similar pattern was found when we had considered Independent and Exchangeable correlation structures into the model. However, the effect of slopes was not statistically significant at 5% level of significance when independent and exchangeable correlation structures considered in the model.

Table 3 depicts the parameter estimations under CIC criteria for all three working correlations structures. When we considered AR (1) correlation structure into the model, we found that the estimated difference in slopes, 0.22, is not significant at 5% level of significance. And also the effect of slopes was not statistically significant at 5% level of significance when independent and exchangeable correlation structures considered in the model. The effect of therapy alone had the similar effect when all three working correlation structures considered simultaneously in the model. the effect was also same between QIC and CIC among these three working

correlation structures. The effect of the visit had similar trend while comparing CIC and QIC among three working correlation structures.

Figure 1 describes Kernel density estimates for distributions of the first component of QIC values when working correlation structure considered as "exchangeable". It is clear that the second set of observations (M=2) has most highest kernel density as compared to first (M=1) and third set (M=3) of observations. Figure 2 represents a distribution of Kernel density estimates of the second component of QIC values when working correlation structure considered as "independence". It is clear that the first set of observations (M=1) has lowest kernel density and third set (M=3) of observations has the highest density. When we have plotted the kernel density estimates for the difference between first (T 1) and the second component (T2) after taking into consideration of independence working correlation structure (Figure 3), it is found that all three sets of observations pretend to have similar kind of distribution. In all three Figures (1-3) the kernel density plot signifies that data follows the normal distribution for three set of observations. Figure 4 provides the bar Diagram for p-values obtained after considering different working correlation structures in the GEE model for QIC & CIC.

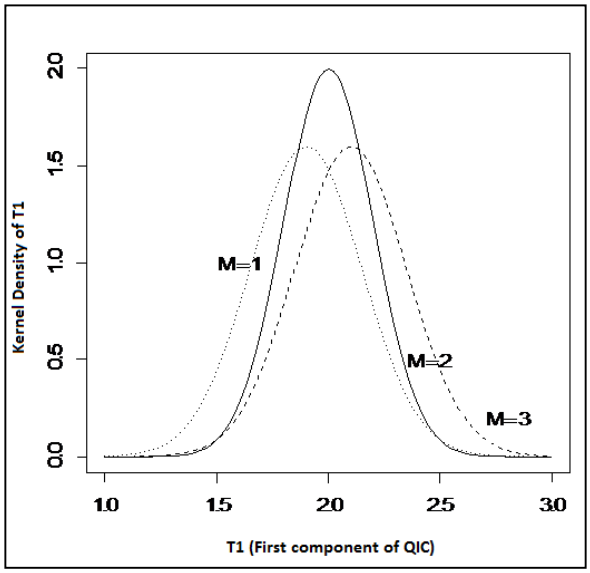

FIGURE 1:-Kernel density estimates of the distributions of first component of QIC values when working correlation structure considered as "exchangeable".

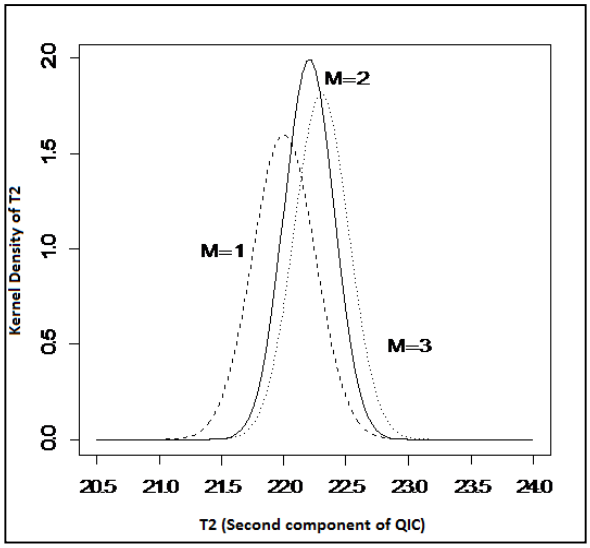

FIGURE 2:-Kernel density estimates of the distributions of second component of QIC values when working correlation structure considered as "independence"

DISCUSSION AND CONCLUSION

The MELD score is considered as the response of interest to compare the effect of a drug on liver cirrhosis patients. The GEE model has been performed to compare QIC and CIC after considering three working correlation structure. In a case of AR(1) correlation structure, it is found that similar regression parameter estimates are observed for both information criteria techniques. However, the effect of therapy over follow-up visit is statistically significant for QIC and it shows that therapy is significantly associated with MELD score and are effective over the period of time. When independent and exchangeable working correlations are considered into GEE model, it is found that effect of therapy over follow-up visit is not statistically significant for QIC and CIC.

The MELD score is applied in the management of patients with chronic liver disease for the non-

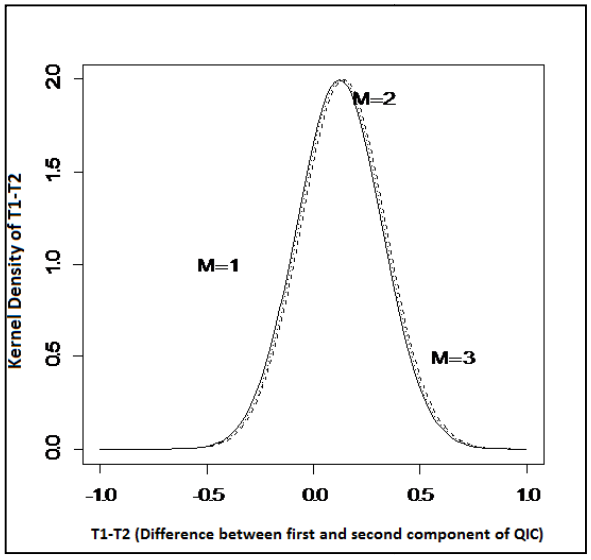

FIGURE 3: Kernel density estimates of the distributions of difference between first and second component of QIC values when working correlation structure considered as "independence".

transplant setting.[14,18] The MELD score is found the useful tool to evaluate the liver cirrhosis disease.[15,19] Although, it is based on three objective laboratory variables, it can be influenced by other clinical variables based on situations. It's also useful for management of patients with a wide spectrum of liver disease. The MELD score is applicable as working models and it served as an outline for further improvement to achieve the goal of equitable distribution of a scare resource. There are several types of suitable model to apply to the repeatedly measured correlated data. It is not possible to detect the perfect model. The approach is to search the most suitable model among selected models. A different model generation is being driven by several correlated structures. Generalized estimating equations are attractive for several reasons, including their relative simplicity.[9] They can include any kind of response distribution among the exponential family.[21] They are hence promising for liver cirrhosis data that are longitudinal. However, the QIC performed not well in our study. Therefore, we cannot advise this information criterion. Consequently, GEEs should only be used when the biological rationale for selecting the covariance structure is obvious (see also a qualitative

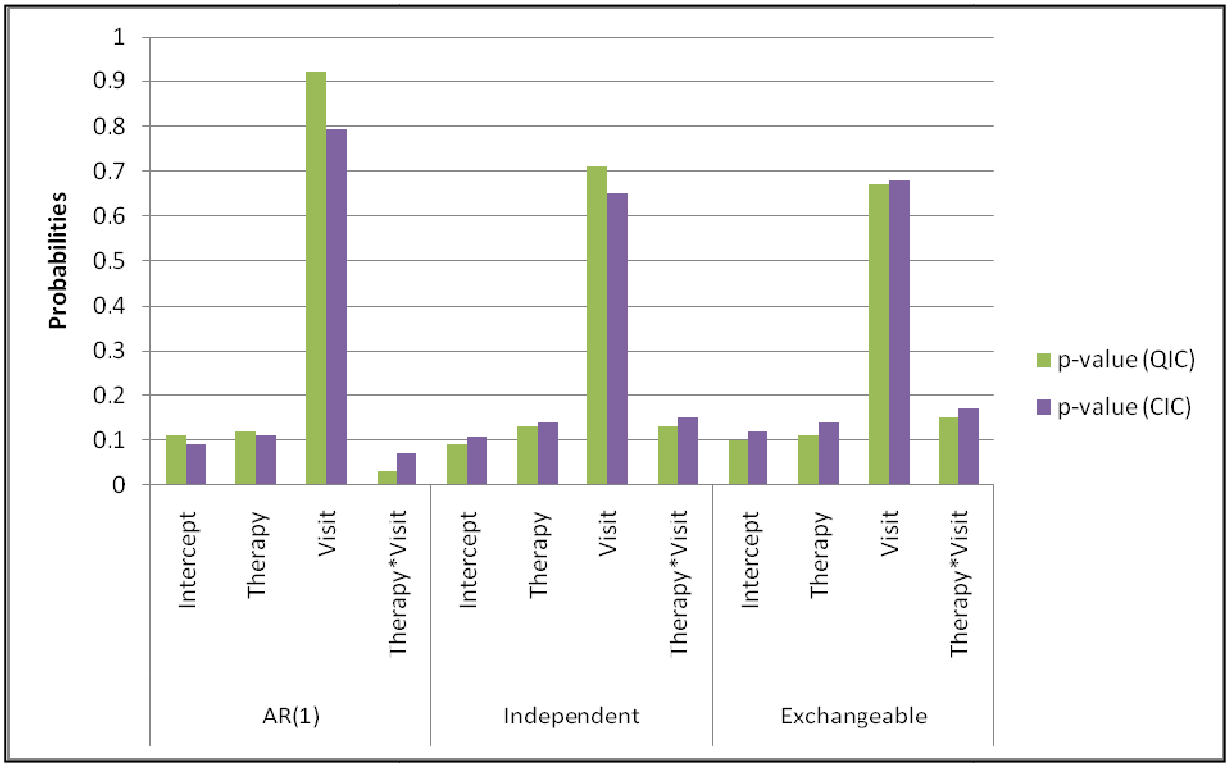

FIGURE 4: Bar Diagram for P-values obtained after considering different working correlation structures in the model for QIC& CIC.

comparison that can be considered.[22] In this framework, it can be stated that the CIC is useful for selecting appropriate correlation structures.

Acknowledgment

The second author acknowledges the support of UGC-BSR one-time grant (F-19-145/2015 (BSR)) of university grant commission, New Delhi.

REFERENCES

1. Bravo AA, Sheth SG, Chopra S. Liver biopsy. N Engl J Med 2001;344(7):495-500.
2. Nath DC, Bhattacharjee A, Vishwakarma RK. Cure rate modeling: -an application with bayesian approach in liver cirrhosis patients. Journal of Data Science 2015;13:421-42.
3. Linhart H, Zucchini W. Model Selection. 1st ed. New York: Wiley; 1986. p.301.
4. Nath DC, Bhattacharjee A. Effect of Correlation Structure in Generalized Estimating Equations and Quasi Least Squares: An Application in Type 2 Diabetes Patients. Pak J Stat Oper Res 2011;7(2):401-12.
5. Wang YG, Carey V. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. Biometrika 2003;90:29-41.
6. Cui J, Qian G. Selection of working correlation structure and best model in GEE analyses of longitudinal data. Communications in Statistics - Simulation and Computation 2007;36(5):987-96.
7. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;57(1):120-5.
8. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22.
9. Lipsitz SR, Fitzmaurice GM, Ibrahim JG, Sinha D, Parzen M, Lipshultz S. Joint generalized estimating equations for multivariate longitudinal binary outcomes with missing data: An application to AIDS data. J R Stat Soc Ser A Stat Soc 2009;172(1):3-20.
10. Vaida F, Blanchard S. Conditional Akaike information for mixed-effects models. Biometrika 2005;92(2):351-70.
11. Hodges J, Sargent D. Counting degrees of freedom in hierarchical and other richly-parameterised models. Biometrika 2001;88:367-79.
12. Ye J. On measuring and correcting the effects of data mining and model selection. J Am Statist Assoc 1998;93(441): 120-31.
13. Kim HJ, Lee HW. Important predictor of mortality in patients with end-stage liver disease. Clin Mol Hepatol 2013;19(2):105-15.
14. Kamath PS, Kim WR. The model for end-stage liver disease (MELD). Hepatology 2007;45(3):797-805.
15. Freeman RB Jr. Model for end-stage liver disease (MELD) for liver allocation: a 5-year score card. Hepatology 2008;47(3): 1052-7.
16. Lake JR. MELD--an imperfect, but thus far the best, solution to the problem of organ allocation. J Gastrointestin Liver Dis 2008;17(1):5-7.
17. Fitzmaurice GM. A caveat concerning independence estimating equations with multivariate binary data. Biometrics 1995;51(1):309-17.
18. Pinzani M, Rombouts K. Liver fibrosis: from the bench to clinical targets. Dig Liver Dis 2004;36(4):231-42.
19. Lau T, Jawad A. Clinical applications of the Model for End-Stage Liver Disease (MELD) in hepatic medicine. Hepat Med 2013;5:1-10.
20. Hin LY, Wang YG. Working-correlation-structure identification in generalized estimating equations. Stat Med 2009;28(4):642-58.
21. Zorn CJW. Generalized estimating equation models for correlated data: a review with applications. Am J Polit Sci 2001;45(2):470-90.
22. Bishop J, Die D, Wang YG. A generalized estimating equations, approach for analysis of the impact of new technology on a trawl fishery. Aust N Z J Statist 2000;42(2):159-77.
23. Nath DC, Bhattacharjee A, Vishwakarma RK. Risk assessment in liver transplantation patients: an shared frailty parametric approach. Clinical Epidemiology and Global Health 2016;4(1):1-15.