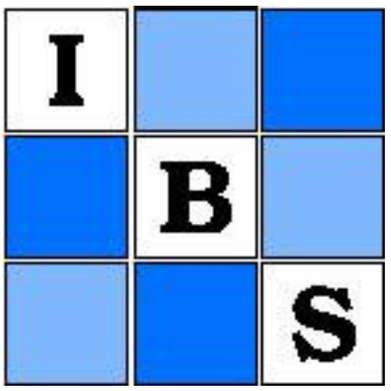

BIOMETRICS 57, 120–125
March 2001

Akaike's Information Criterion in Generalized Estimating Equations

Wei Pan

Division of Biostatistics, University of Minnesota,
MMC 303, 420 Delaware Street SE, Minneapolis, Minnesota 55455, U.S.A.
email: weip@biostat.umn.edu

SUMMARY. Correlated response data are common in biomedical studies. Regression analysis based on the generalized estimating equations (GEE) is an increasingly important method for such data. However, there seem to be few model-selection criteria available in GEE. The well-known Akaike Information Criterion (AIC) cannot be directly applied since AIC is based on maximum likelihood estimation while GEE is nonlikelihood based. We propose a modification to AIC, where the likelihood is replaced by the quasi-likelihood and a proper adjustment is made for the penalty term. Its performance is investigated through simulation studies. For illustration, the method is applied to a real data set.

KEY WORDS: Akaike Information Criterion; Generalized estimating equations; Generalized linear models; Model selection; Quasi-likelihood.

1. Introduction

Correlated response data arise often from biomedical studies. An example to be studied is the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) (Klein et al., 1984), where a binary response variable is the presence of diabetic retinopathy in each of the two eyes from each participant in the study. Since the two observations on the two eyes from the same participant tend to be correlated, statistical analyses have to take proper account of this correlation. Since the publication of the seminal paper by Liang and Zeger (1986), the generalized estimating equation (GEE) approach has become increasingly important in handling such correlated data.

Model selection is an important issue in almost any practical data analysis. A common problem is variable selection in regression: given a large group of covariates (including some higher order terms), one needs to select a subset to be included in the regression model. In the WESDR, 13 potential risk factors were collected, and we need to determine which of these factors are to be included. It is well known that, in observational studies such as the WESDR, excluding some important risk factors (i.e., confounders) may result in misleading estimates of the effects of other risk factors. On the other hand, including all covariates may lead to a too complex model with difficulty in interpretation and with less precise parameter estimates.

There is an extensive model-selection literature in statistics (e.g., Miller, 1990, and references therein) but mainly for the classic linear regression with independent data. One powerful and widely used model-selection criterion is Akaike's Information Criterion (AIC) (Akaike, 1973). AIC is based on the likelihood and asymptotic properties of the maximum likelihood estimator (MLE). Since no distribution is assumed in generalized estimating equations (GEE), there is no likelihood defined; thus, AIC cannot be directly used. On the other hand, the issue of model selection in GEE has been largely neglected. The goal of this article is to propose an extension of AIC to GEE. It involves using the quasi-likelihood constructed from the estimating equations (Wedderburn, 1974). Since in general the GEE estimator has different asymptotic properties from those of the MLE, a modification to the penalty term in the usual AIC is also necessary.

This article is organized as follows. In Section 2, we first briefly review the GEE and quasi-likelihood; then we propose a modification to AIC in GEE. Simulation results are presented in Section 3 to show its performance in selecting the working correlation matrix and selecting covariates in GEE. Section 4 applies the method to the WESDR data, followed by a brief discussion.

2. AIC in GEE

2.1 *GEE*

Suppose we have a random sample of observations from n individuals. For each individual i, we have a vector of responses $Y_i = (Y_{i1}, \ldots, Y_{in_i})'$ and corresponding covariates $X_i = (X'_{i1}, \ldots, X'_{in_i})'$, where each Y_{ij} is a scalar and X'_{ij} is a p-vector. In general, the components of Y_i are correlated but Y_i and Y_k are independent for any $i \neq k$ (conditional on the covariates). We use $\mathcal{D} = \{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ to denote the data at hand. To model the relation between the response and covariates, one can use a regression model similar to the generalized linear models, $g(\mu_i) = X_i\beta$, where $\mu_i = \mathrm{E}(Y_i \mid X_i)$, g is a specified link function, and $\beta = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown regression coefficients to be estimated. The GEE approach estimates β through solving the following

estimating equations (Liang and Zeger, 1986):

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

where $D_i = D_i(\beta) = \partial\mu_i(\beta)/\partial\beta'$ and V_i is a working covariance matrix of Y_i. V_i can be expressed in terms of a working correlation matrix $R = R(\alpha)$, $V_i = A_i^{1/2} R(\alpha) A_i^{1/2}$, where A_i is a diagonal matrix with elements $\text{var}(Y_{ij}) = \phi V(\mu_{ij})$, which is specified as a function of the mean μ_{ij}. The α may be some unknown parameters involved in the working correlation structure, which can be estimated through the method of moments or another set of estimating equations.

An attractive point of the GEE approach is that it yields a consistent estimator of β, $\hat{\beta}$, even when the working correlation matrix R is misspecified (Liang and Zeger, 1986). For instance, it is often convenient to use a working independence model where $R = I$. Some other popular choices include compound symmetry (CS) (i.e., exchangeable) with $R_{ij} = \rho$ for any $i \neq j$ or first-order autoregressive (AR-1) with $R_{ij} = \rho^{|i-j|}$, where R_{ij} denotes the (i, j)th element of R. Due to its simplicity, the working independence model is attractive. Many studies have shown that $\hat{\beta}$ obtained under the independence model is relatively efficient (Zeger, 1988; McDonald, 1993), at least when the correlation between responses is not large. Another compelling reason for using the working independence model is in partly conditional modeling of means for longitudinal data (Pepe and Anderson, 1994). However, for time-varying or cluster-specific covariates, Fitzmaurice (1995) showed that the resulting estimator from the independence model may be very inefficient; its efficiency may be as low as 60% compared with the estimator obtained by using the correct correlation structure. Hence, this poses a model-selection problem in selecting the working correlation structure. Of course, we may also need to decide which covariates are to be included in the regression model $g(\mu_i)$. Below we propose a quasi–likelihood-based model-selection criterion that can be applied to address the above issues.

2.2 *Quasi-Likelihood*

Now we need to briefly review the quasi-likelihood. For the moment, suppose we only have a scalar response variable, y. We first construct the quasi-likelihood function for the mean parameter $\mu = \text{E}(y)$ (and dispersion parameter ϕ); then we will write it in terms of the regression parameter β.

Based on the model specification $\text{E}(y) = \mu$ and $\text{var}(y) = \phi V(\mu)$, the (log) quasi-likelihood function is (McCullagh and Nelder, 1989, p. 325)

$$Q(\mu, \phi; y) = \int_y^{\mu} \frac{y - t}{\phi V(t)} dt. \tag{2}$$

For instance, with grouped binary data, $y \sim \text{Bin}(n, \pi)$ it is often specified that $V(\mu) = \mu(1 - \mu/n)$; then (up to a constant) $Q(\mu, \phi; y) = L(\mu, \phi; y)/\phi$, where $L(\mu, \phi; y) = y \times \log[\mu/(n - \mu)] + n \log(n - \mu)$ is the log likelihood for the binomial distribution. When $\phi = 1$, the quasi-likelihood Q reduces to L. However, $\phi > 1$ is extremely useful in modeling overdispersion that commonly occurs in practice. Some common examples of the quasi-likelihood are given in McCullagh and Nelder (1989, p. 326).

With a $1 \times p$ covariate x and a specified regression model $\text{E}(y) = \mu = g^{-1}(x\beta)$ and $\text{var}(y) = \phi V(\mu)$, the quasi-likelihood can be written as a function of the regression coefficients β, i.e., $Q(\beta, \phi; (y, x)) = Q(g^{-1}(x\beta), \phi; y)$.

In the current context, if the working independence model $R = I$ is used, the working assumption is that the paired observations (Y_{ij}, X_{ij}) in $\mathcal{D}$ are independent. Hence, the quasi-likelihood based on $\mathcal{D}$ is

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

It is easy to verify that the left-hand side of the GEE $S(\beta; I, \mathcal{D})$ in (1) is equivalent to $\partial Q(\beta, \phi; I, \mathcal{D})/\partial\beta$. Thus, the GEE (1) can be regarded as a quasi-likelihood score equation.

However, if we use a more general working correlation matrix R, there is no guarantee that a corresponding quasi-likelihood exists unless certain conditions are satisfied (McCullagh and Nelder, 1989, p. 333–335). Furthermore, even if it exists, in general it is difficult to construct. How to construct a quasi-likelihood with a general working correlation matrix is beyond the scope of this article. The main goal of this article is to propose a criterion based on $Q(\hat{\beta}, \phi; I, \mathcal{D})$, the quasi-likelihood under the working independence model with an estimated β, using any general working correlation structure in GEE.

2.3 *AIC and a Modification to AIC in GEE*

We first briefly review the derivation of AIC, which will motivate our modification to AIC. A more rigorous and general discussion is available from Linhart and Zucchini (1986). For simplicity of notation, we first assume that the dispersion parameter ϕ is known; hence, we can ignore it in the (quasi-)likelihood function. At the end of this section, we will discuss the situation when ϕ is unknown.

Suppose we have a candidate model M_1 and the true model M_* with log-likelihood functions $L(\beta; \mathcal{D})$ and $L(\beta_*; \mathcal{D})$, respectively. Throughout, we assume that each model can be indexed by the parameter vector β. A well-known measure of separation between two models is given by the Kullback–Leibler information (Kullback and Leibler, 1951), also known as the cross entropy. The Kullback–Leibler information between M_1 and M_* is

$$\Delta_0(\beta, \beta_*) = \text{E}_{M_*}[-2L(\beta; \mathcal{D})], \tag{3}$$

where the expectation E_{M_*} is taken with respect to the true distribution of $\mathcal{D}$ (i.e., under model M_*). From a set of candidate models $\mathcal{M}$, in which each can be indexed by β, we would like to choose the model with the smallest $\Delta_0(\beta, \beta_*)$. However, in practice, since both β and β_* are unknown, we have to estimate $\Delta_0(\beta, \beta_*)$. AIC was motivated as an asymptotically unbiased estimator of $\text{E}_{M_*}[\Delta_0(\hat{\beta}, \beta_*)]$, where $\hat{\beta}$ is the maximum likelihood estimator (MLE) under any candidate model in $\mathcal{M}$ and the expectation is taken over the random $\hat{\beta}$. Akaike proposed using AIC as a model-selection criterion, i.e.,

$$AIC = -2L(\hat{\beta}; \mathcal{D}) + 2p, \tag{4}$$

where p is the dimension of β. Model selection is accomplished by selecting from $\mathcal{M}$ the one that minimizes AIC.

Since GEE is nonlikelihood based, we do not have a likelihood function in this context. However, we may have a quasi-

likelihood. We propose replacing the likelihood L in (3) by the quasi-likelihood Q under the working independence model and define a new discrepancy as

$$\Delta(\beta, \beta_*, I) = \mathrm{E}_{M_*}[-2Q(\beta; I, \mathcal{D})]. \tag{5}$$

We assume that any quasi-likelihood model in $\mathcal{M}$ can be indexed by the parameter vector β and that β_* is the corresponding parameter for the quasi-likelihood model induced by the true data-generating model M_*. For simplicity, with a slight abuse of notation, we suppress the dependence of $\Delta(\beta, \beta_*, I)$ on the true model M_*. It is well known that

$$\mathrm{E}_{M_*}\left(-\frac{\partial Q(\beta; I, \mathcal{D})}{\partial \beta}\bigg|_{\beta=\beta_*}\right) = 0,$$

$$\Omega_I = \mathrm{E}_{M_*}\left(-\frac{\partial^2 Q(\beta; I, \mathcal{D})}{\partial\beta\partial\beta'}\bigg|_{\beta=\beta_*}\right) = \sum_{i=1}^{n} D_i' V_i D_i,$$

and the latter is positive semidefinite. Under suitable conditions, one can exchange the order of the integration and differentiation. Then β_* is a local minimizer of $\Delta(\beta, \beta_*, I)$ with regard to β. In other words, for any β in a neighborhood of β_*, we have

$$\Delta(\beta, \beta_*, I) \geq \Delta(\beta_*, \beta_*, I). \tag{6}$$

This implies that the discrepancy $\Delta(\beta, \beta_*, I)$ is well defined for all the models close to the true model. Though we cannot prove β_* is in general a global minimizer of $\Delta(\beta, \beta_*, I)$, in the common situation that the marginal quasi-likelihood $Q(\beta; (Y_{ij}, X_{ij}))$ is equal to the log likelihood $L(\beta; (Y_{ij}, X_{ij}))$, it is straightforward to verify that then β_* is indeed a global minimizer of $\Delta(\beta, \beta_*, I)$ due to the fact that $\mathrm{E}_{M_*}[L(\beta_*; (Y_{ij}, X_{ij}))] > \mathrm{E}_{M_*}[L(\beta; (Y_{ij}, X_{ij}))]$ for any $\beta \neq \beta_*$ (cf., Lehmann, 1983, p. 409).

Now suppose the GEE estimator $\hat{\beta} = \hat{\beta}(R)$ is obtained using any general working correlation structure R. Following the idea of deriving Proposition 2 of Linhart and Zucchini (1986, p. 241, which is for minimum discrepancy estimators), we can approximate $\mathrm{E}_{M_*}[\Delta(\hat{\beta}, \beta_*, I)]$ as

$$\begin{aligned}\mathrm{E}_{M_*}[\Delta(\hat{\beta}, \beta_*, I)] \approx & -2\mathrm{E}_{M_*}[Q(\hat{\beta}; I, \mathcal{D})] \\ & + 2\mathrm{E}_{M_*}[(\hat{\beta} - \beta_*)' S(\hat{\beta}; I, \mathcal{D})] \\ & + 2\,\mathrm{trace}(\Omega_I J),\end{aligned} \tag{7}$$

where $J = \mathrm{cov}(\hat{\beta})$, which can be consistently estimated by the robust or sandwich covariance estimator, say, $\hat{V}_r$ (Liang and Zeger, 1986). Ω_I can also be consistently estimated by its empirical estimator $\hat{\Omega}_I = -\partial^2 Q(\beta; I, \mathcal{D})/\partial\beta\partial\beta'|_{\beta=\hat{\beta}}$. Note that, for $\hat{\beta} = \hat{\beta}(R)$, we have $S(\hat{\beta}; R, \mathcal{D}) = 0$ but not necessarily $S(\hat{\beta}; I, \mathcal{D}) = 0$ unless $R = I$. By ignoring the second term that is difficult to estimate, we have an estimator of the right-hand side of (7),

$$QIC(R) \equiv -2Q(\hat{\beta}(R); I, \mathcal{D}) + 2\,\mathrm{trace}(\hat{\Omega}_I \hat{V}_r). \tag{8}$$

This is our proposed quasi-likelihood under the independence model criterion (QIC) for GEE. Our simulation results (see Section 3) show that ignoring the second term in (7) does not dramatically, but does somewhat, influence the performance of $QIC(R)$, and $QIC(I)$ is the best. Note that, if the working independence model is used in GEE, by the consistency of $\hat{\beta}$, $\hat{\Omega}_I$, and $\hat{V}_r$ and that $S(\hat{\beta}; I, \mathcal{D}) = 0$, we know $QIC(I)$ is an asymptotically unbiased estimator of (7). Furthermore, $\hat{\Omega}_I$ and $\hat{V}_r$ are directly available from the model fitting results in many statistical packages, such as SAS and S-Plus. Hence, we recommend the routine use of $QIC(I)$ whenever possible. QIC can also be applied to select a working correlation structure in GEE: one needs to calculate the QIC for various candidate working correlation structures and then pick the one with the smallest QIC. Note that here the goal of selecting a working correlation structure is to estimate β more efficiently.

In practice, since ϕ is unknown, we plug in $\hat{\phi}$, which is estimated from the largest model available. In variable selection, that means we estimate ϕ based on the regression model including all covariates. This is similar to estimating the dispersion parameter in linear regression with Mallows' (1973) C_p. A more general but also more difficult approach is to use the extended quasi-likelihood (McCullagh and Nelder, 1989, p. 349), which we do not pursue here.

2.4 *Remarks*

When all modeling specifications in GEE are correct, $\hat{\Omega}_I^{-1}$ and $\hat{V}_r$ are asymptotically equivalent and $trace(\hat{\Omega}_I \hat{V}_r) \approx trace(I) = p$. Then QIC reduces to AIC. In GEE with correlated data, one may take $QIC_u(R) \equiv -2Q(\hat{\beta}(R); I, \mathcal{D}) + 2p$ as an approximation to $QIC(R)$, and thus $QIC_u(R)$ can be potentially useful in variable selection. However, it is easy to see that $QIC_u(R)$ cannot be applied to select the working correlation matrix R.

Our main motivation of defining the discrepancy $\Delta(\beta, \beta_*, I)$ using $Q(\beta; I, \mathcal{D})$ is the latter's simplicity and uniqueness. However, as suggested by one referee, it may be possible to define a more general discrepancy as $\Delta(\beta, \beta_*, R) = \mathrm{E}_{M_*}[-2Q(\beta; R, \mathcal{D})]$. But note that $Q(\beta; R, \mathcal{D})$ may not be unique and in general can be calculated as a path-dependent line integral (McCullagh and Nelder, 1989, Section 9.3.2). Nevertheless, according to Theorem 1 of Hanfelt and Liang (1995; see also Li, 1993), $\Delta(\beta, \beta_*, R)$ is still a well-defined discrepancy in the sense of (6).

3. Simulations

Simulation studies were conducted to investigate the performance of our proposed model-selection criterion QIC in selecting the working correlation structure and selecting the covariates in a marginal logistic regression model. We used the same true model as in Fitzmaurice (1995). The response variable Y_{it} is binary and its marginal mean is μ_{it}, with

$$\mathrm{logit}(\mu_{it}) = \beta_0 + \beta_1 x_{1,it} + \beta_2(t-1), \qquad \begin{array}{l} t = 1, 2, 3 \text{ and} \\ i = 1, \ldots, n, \end{array}$$

where the $x_{1,it}$ are i.i.d. Bernoulli, i.e., $x_{1,it} = 0$ or 1 with probability 1/2 and $\beta_0 = 0.25 = -\beta_1 = -\beta_2$. The true correlation matrix is CS. We used a large correlation, $\rho = 0.5$, and moderate sample size, $n = 50$ or 100. The joint distribution of the Y_i was simulated from Bahadur's (1961) representation (see Fitzmaurice, 1995, for more details).

For each sample size, $n = 50$ or 100, our proposed method is most likely to correctly select the CS from the three given correlation structures (Table 1). Since the distribution form of the data is known, we can also compute the MLE and thus AIC. For comparison, we also attach the results of using AIC by assuming various correlation matrices. Unsurprisingly,

Table 1

Frequency of the working correlation matrix selected by QIC versus AIC for the marginal logistic model from 1000 *independent replications. The true correlation matrix is CS.*

	$n = 50$			$n = 100$		
Criterion	Ind	CS	AR-1	Ind	CS	AR-1
QIC	138	678	184	140	721	139
AIC	0	836	164	0	946	54

AIC is more efficient than is QIC, probably for two reasons. First, the MLE of β is more efficient than the GEE estimator. Second, information on the true correlation structure is embedded in the likelihood function in AIC but not directly in the quasi-likelihood $Q(\beta; I, \mathcal{D})$ in QIC. As mentioned earlier, the strength of QIC is that it is nonlikelihood based, whereas in practice the likelihood approach is often too restrictive with its strong distributional assumption for correlated categorical data.

Now we consider variable selection with an expanded full model,

$$\text{logit}(\mu_{it}) = \beta_0 + \beta_1 x_{1,it} + \beta_2(t-1) + \beta_3 x_{3,it} + \beta_4 x_{4,it}, \quad t = 1, 2, 3 \text{ and } i = 1, \ldots, n,$$

where $x_{1,it}$, β_0, β_1, and β_2 are as before, $x_{3,it}$ and $x_{4,it}$ are i.i.d. uniform $\text{U}(-1, 1)$ and independent of $x_{1,it}$, and $\beta_3 = \beta_4 = 0$. For simplicity, we consider five nonnested candidate models with various subsets of covariates included. The results of using QIC with different working correlation matrices are shown in Table 2. The performance of the three QICs with different working correlation matrices is close, but QIC(Ind) appears to be the best. This is probably related to the error introduced by ignoring the second term in (7) for QIC(CS) and QIC(AR-1). For comparison, we also list the results of using AIC under the correct and incorrect correlation structures. Surprisingly, QIC(Ind) turns out to be comparable with AIC/CS. When the distributional assumptions are violated, the performance of AIC deteriorates, as demonstrated by AIC/Ind and AIC/AR-1, which incorrectly assume the independence and AR-1 correlation matrices, respectively.

We also did simulation studies to investigate the QICs performance in selecting the working correlation matrix in modeling a partly conditional mean for longitudinal data (Pepe and Anderson, 1994) and in variable selection for correlated overdispersed (grouped) binary data. The results (not shown here) also appeared to be promising.

4. An Example

We apply the method to the WESDR (Klein et al., 1984). The study goal was to determine the risk factors for diabetic retinopathy. The binary response is the presence of diabetic retinopathy in each of two eyes from each of 720 individuals in the study. There are 13 potential risk factors. As shown in Barnhart and Williamson (1998), a univariate analysis was conducted to investigate the marginal association between the response variable and each risk factor. It was found that eight of them are marginally associated with the response variable. Barnhart and Williamson included only four risk factors, i.e., duration of diabetes (years), glycosylated hemoglobin level, diastolic blood pressure, and body mass index, plus the two quadratic terms of duration of diabetes and body mass index in their final model. Now we consider adding all or some of the four removed covariates (i.e., intraocular pressure, systolic blood pressure, pulse rate, and proteinuria) into Barnhart and Williamson's model. Hence, we have 16 candidate models. Note that these models cannot be ordered as a nested sequence, and one advantage of using a flexible model-selection criterion such as QIC is its ability to compare nonnested models. Due to the nature of the possible correlation between the two observations on the two eyes from the same participant, GEE is used to fit the marginal logistic regression model and QIC is applied to do model selection, all under the working independence model. The selected top four models, along with the full model (ranked 8) and Barnhart and Williamson's model (ranked 10), are listed in Table 3. The p-values associated with GEE estimates are also presented. According to the QIC values, the top four models are very close but different from Barnhart and Williamson's model in that proteinuria is included in the former four models. From Table 3, we can see that proteinuria is an important (and statistically significant) risk factor, and adding intraocular pressure or systolic blood pressure into the model may also improve its performance.

5. Discussion

For likelihood-based methods, there are many well-studied model-selection criteria, such as AIC. But for nonlikelihood-based methods, such as GEE, there is a lack of literature on model selection. In this article, we have proposed a new criterion QIC that works for GEE. The QIC involves using

Table 2

Frequency of the set of variables selected by QIC versus AIC for the marginal logistic model from 1000 *independent replications. The true model has* $\{x_1, x_2\}$*, and AIC/CS is calculated correctly using the CS correlation matrix.*

	$n = 50$					$n = 100$				
Criterion	x_1	x_1, x_2	x_1, x_3	x_1, x_2, x_3	x_1, x_2, x_3, x_4	x_1	x_1, x_2	x_1, x_3	x_1, x_2, x_3	x_1, x_2, x_3, x_4
QIC(Ind)	272	488	69	80	91	108	657	31	121	83
QIC(CS)	236	444	126	108	86	95	618	53	143	91
QIC(AR-1)	249	450	124	01	76	97	636	55	136	76
AIC/Ind	493	304	83	51	69	236	549	54	93	68
AIC/CS	240	441	53	123	143	112	631	14	135	108
AIC/AR-1	307	322	78	110	183	151	535	39	132	143

Table 3
QIC and robust p-values for each covariate in the top four models and the other two models with the WESDR data

Covariate	Model					
	1	2	3	4	8	10
Intraocular pressure	0.1411	—	0.1378	—	0.1487	—
Systolic blood pressure	—	—	0.1982	0.1982	0.1821	—
Pulse rate	—	—	—	—	0.1991	—
Proteinuria	0.0412	0.0375	0.0356	0.0321	0.0397	—
Duration of diabetes	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Glycosylated hemoglobin	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Diastolic blood pressure	0.0403	0.0187	0.0106	0.0045	0.0282	0.0044
Body mass index	0.0001	0.0001	<0.0001	<0.0001	<0.0001	<0.0001
$(\text{Duration of diabetes})^2$	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
$(\text{Body mass index})^2$	0.0001	0.0001	0.0001	0.0001	<0.0001	0.0001
QIC(Ind)	1185.5	1185.7	1185.8	1186.0	1186.5	1189.8

the quasi-likelihood constructed under the working independence model and the naive and robust covariance estimates of estimated regression coefficients. Although using other more general quasi-likelihood seems possible, we choose to use the quasi-likelihood under the working independence model due to its simplicity. However, QIC allows one to use any general working correlation structure to estimate the parameters in GEE. In simulation studies, we found that the QIC works well in variable selection and selecting the working correlation matrix. We were particularly impressed with the performance of QIC(I) in variable selection. Further applications warrant future studies.

Acknowledgement

The author thanks Dr Huiman Barnhart for providing the WESDR data set. The author is grateful to Dr Lynn Eberly, two referees, an associate editor, and the editor for extremely thorough and helpful comments that greatly improved the article.

Résumé

Les données à réponses corrélées sont habituelles dans les études biomédicales. L'analyse de régression basée sur les équations d'estimation généralisées (GEE) est une méthode d'importance croissante pour de telles données. Poutant, il semble exister peu de critères de sélection de modèles disponibles pour GEE. Le critère d'information d'Akaike (AIC) bien connu, ne peut être appliqué directement, étant donné que l'AIC est basé sur l'estimation du maximum de vraisemblance, alors que GEE est basé sur la quasi-vraisemblance. Nous proposons une modification de AIC, où la vraisemblance est remplacée par la quasi-vraisemblance et un ajustement adapté est fait pour le terme de pénalité. Ses performances sont évaluées au travers d'études de simulation. Pour illustration, la méthode est appliquée à un jeu de données réel.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory,* B. N. Petrov and F. Csaki (eds), 267–281. Budapest: Akademiai Kiado.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction,* Volume VI, *Stanford Mathematical Studies in the Social Sciences,* H. Solomon (ed.), 158–168. Stanford, California: Stanford University Press.

Barnhart, H. X. and Williamson, J. M. (1998). Goodness-of-fit tests for GEE modeling with binary data. *Biometrics* **54,** 720–729.

Fitzmaurice, G. M. (1995). A caveat concerning independence estimating equations with multiple multivariate binary data. *Biometrics* **51,** 309–317.

Hanfelt, J. J. and Liang, K.-Y. (1995). Approximate likelihood ratios for general estimating functions. *Biometrika* **82,** 461–477.

Klein, R., Klein, B. E. K., Moss, S. E., Davis, M. D., and DeMets, D. L. (1984). The Wisconsin Epidemiologic Study of Diabetic Retinopathy: II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. *Archives of Ophthalmology* **102,** 520–526.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* **22,** 79–86.

Lehmann, E. L. (1983). *Theory of Point Estimation.* New York: Wiley.

Li, B. (1993). A deviance function for the quasi-likelihood method. *Biometrika* **80,** 741–753.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Linhart, L. and Zucchini, W. (1986). *Model Selection.* New York: Wiley.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15,** 661–675.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd edition. London: Chapman and Hall.

McDonald, B. W. (1993). Estimating logistic regression parameters for bivariate binary data. *Journal of the Royal Statistical Society, Series B* **55,** 391–397.

Miller, A. J. (1990). *Subset Selection in Regression.* London: Chapman and Hall.

Pepe, M.S. and Anderson, G. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Series B* **23,** 939–951.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61,** 439–447.

Zeger, S. L. (1988). The analysis of discrete longitudinal data: Commentary. *Statistics in Medicine* **7,** 161–168.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **42,** 121–130.

Received June 1999. *Revised December* 1999 *and June* 2000. *Accepted June* 2000.