이 학 박 사 학 위 논 문

H-likelihood approach for clinical pharmacology data

다단계 우도를 이용한 임상 약리 자료 분석

2020년 2월

서울대학교 대학원

통계학과

안 형 미

이 학 박 사 학 위 논 문

H-likelihood approach for clinical pharmacology data

다단계 우도를 이용한 임상 약리 자료 분석

2020년 2월

서울대학교 대학원

통계학과

안 형 미

Ph.D. DISSERTATION

H-likelihood approach for clinical pharmacology data

by

Hyungmi An

A Thesis
submitted in fulfillment of the requirement
for the degree of
Doctor of Philosophy
in
Statistics

Department of Statistics
College of Natural Science
Seoul National University
February, 2020

Abstract

H-likelihood approach proposed by Lee and Nelder (1996) is widely used for various data. In particular, repeated measured data within clusters can be analyzed by hierarchical generalized linear models (HGLMs). When we are interested in the multiple endpoints which are correlated, then multivariate double hierarchical generalized linear models (multivariate double HGLMs) can be considered.

In this thesis, we apply multivariate double hierarchical generalized linear models for bioequivalence testing which is performed to assess the similarity in the pharmacokinetic profiles between a test product and its reference product. If the 90% confidence interval for the geometric mean ratio (GMR) of a test to the reference product entirely falls within the bioequivalence margin, $(0.8, 1.25)$, for both AUC and C_{max}, the test product is declared to be bioequivalent. Since two co-primary endpoints AUC and C_{max} are strongly correlated, we consider multivariate double HGLMs which provide smaller standard errors of estimated treatment effects and resulted in narrower 90% confidence interval for GMR.

To select the best fitting model among different model classes, we define conditional Akaike information for double hierarchical generalized linear models (double HGLMs) and propose its asymptotically unbiased estimator, conditional Akaike information criterion (cAIC), using effective degree of freedom for double HGLMs. We compare the accuracy and model selection performance of the proposed cAIC with conventional cAIC, and apply it to the real data for the best model selection.

Keywords: h-likelihood, multivariate double hierarchical generalized linear models , bioequivalence tests, effective degree of freedom, model selection, conditional Akaike information criterion

Student Number: 2007-20255

Contents

List of Tables

List of Figures

Chapter 1

Introduction

We deal with h-likielihood approach for chlinical pharmacology data. This thesis is composed of two parts. In chapter 2, we apply multivariate double hierarchical generalized linear models (multivariate double HGLMs) to test the bioequivalence between a test product and its reference product. In chapter 3, we define the conditional Akaike information for double hierarchical generalized linear models (double HGLMs) and propose its asymptotically unbiased estimator conditional Akaike information criterion (cAIC) to select the best fitting model.

In chapter 2, we suggest the multivariate double hierarchical generalized linear models for clinical pharmacology data analysis. Conventional bioequivalence tests are performed using independent linear mixed models for two co-primary endpoints, AUC and C_{max}. However, it is well known that pharmacokinetic outcomes such as AUC and C_{max} are strongly correlated, then the correlations should be incorporated into the model. The class of multivariate double hierarchical generalized linear models is a h-likelihood approach to simultaneously deal with correlated multiple endpoints such as AUC and C_{max}, in which random effects are specified for the dispersion to model the heterogeneity of within-cluster variances. By proper modelling for dispersion and correlations, the gain in information provides smaller standard errors of treatment effects and results in the narrower 90% confidence interval of the geometric mean ratio

(GMR) included into the bioequivalence margin (0.8, 1.25).

In chapeter 3, we study the conditional Akaike information (cAI) for double hierarchical generalized linear models and its estimator, conditional Akaike information criterion (cAIC) to select the best fitting model. Because the effective degree of freedom is closely related to the bias correction term of cAIC, we extend the scope of effective degrees of freedom to cover the dispersion model using h-likelihood. This effective degree of freedom takes account of the uncertainty for estimating parameters in the dispersion model as well as those in the mean model. We propose cAIC using the effective degree of freedom as the bias correction term for double HGLMs, and show that it is an asymptotic unbiased estimator for the cAI. We compare the accuracy and model selection performance of the proposed cAIC to other cAICs using different bias correction terms. Simulation results display the asymptotic unbiasedness and the correct model identification percentageof the proposed cAIC.

Chapter 2

Multivariate double hierarchical generalized linear models for clinical pharmacology data

2.1 Backgrounds

The bioequivalence test is conducted to assess the similarity in the biopharmaceutical performance between a test product and its reference product. The basic idea is that if they are pharmaceutically equivalent (i.e., the active ingredient, formulation, dosage strength, and route of administration are the same) and their pharmacokinetic profiles are sufficiently similar (i.e., bioequivalent), the therapeutic outcome should be the same or the products are therapeutically equivalent (Figure 2.1). The common measures used to show bioequivalence are the area under the concentration-time curve (AUC) and the peak concentration (C_{max}), and they are assumed to be log-normally distributed. If the 90% confidence interval for the geometric mean ratio (GMR) of a test to the reference product entirely falls within the bioequivalence margin, for example, (0.8,1.25) for both AUC and C_{max}, the test product is declared to be bioequivalent. This bioequivalence test has also been frequently adopted in comparative pharmacokinetic studies and drug-drug interaction studies, where similarity in the pharmacokinetic profiles is of interest.

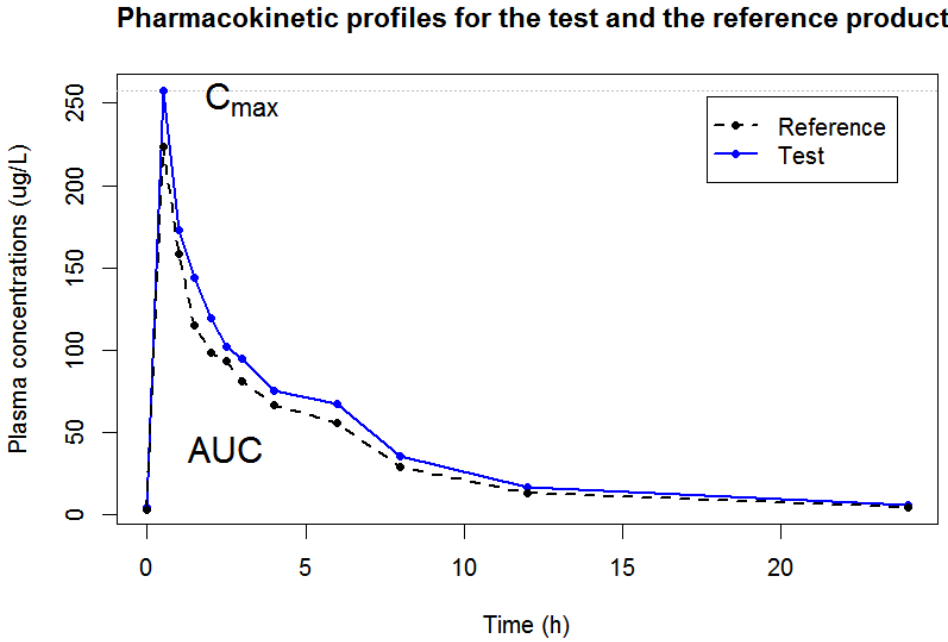

Figure 2.1: Pharmacokinetic profiles for the test product (solid line) and its reference product (dashed line). AUC refers the area under the concentration-time curve, and C_{max} reperesents the peak concentration.

Let y_{ij} be the repeatedly measured outcome from subject i in the period j with a cross-over design ($i = 1, \cdots, m; j = 1, \cdots, q$). Consider a linear mixed model with period, sequence, treatment as fixed effects and subject as random effects,

$$y_{ij} = \beta_0 + \beta_1\, trt_{ij} + \beta_2\, prd_j + \beta_3\, seq_i + v_i + e_{ij} \tag{2.1}$$
$$v_i \sim N(0, \lambda),\ e_{ij} \sim N(0, \phi),$$

where $trt_{ij} = 1$ for the test product and $trt_{ij} = 0$ for the reference product, prd_j and seq_i are covariates indicating period j and sequence of the subject i, and v_i is the subject-specific random effect. The variances λ and ϕ represent between-subject variance and within-subject variance, respectively. The standard error of the treatment effect β_1 determines the width of the 90% confidence interval for geometric mean ratio (GMR) of a test product to the reference product, and it depends largely on the within-subject variance ϕ. The within-subject coefficient of variation CV_w is defined as

$$CV_w = \sqrt{\exp(\phi) - 1}.$$

In the case of $CV_w \geq 0.3$, the drug is defined as 'Highly variable drugs'.

When we fit the model (2.1) for $\log(\text{AUC})$ or $\log(\text{C}_{\text{max}})$ as a responese, then $\exp(\hat{\beta}_1)$, the estimate for the GMR of the test product to the reference, can be obtained for the bioequivalence tesing. If the bioequivalence margin is (0.8, 1.25), then the test problem for bioequivalce is formulated as follows:

$$H_0 : \beta_1 \leq \log(0.8) \text{ or } \beta_1 \geq \log(1.25) \quad v.s. \quad H_1 : \log(0.8) < \beta_1 < \log(1.25),$$

which can be splited into two one-sided tests such that

$$\begin{aligned} & H_{01} : \beta_1 \leq \log(0.8) && v.s. && H_{11} : \beta_1 > \log(0.8), \\ \text{and} \quad & H_{02} : \beta_1 \geq \log(1.25) && v.s. && H_{12} : \beta_1 < \log(1.25). \end{aligned}$$

Then, it can be seen

$$H_0 = H_{01} \cup H_{02} \quad v.s. \quad H_1 = H_{11} \cap H_{12}.$$

According to the intersection-union principle (Berger and Hsu, 1996), H_0 is rejected at significance level 0.05 in favor of H_1 if both null hypotheses H_{01} and H_{02} are rejected at significance level 0.05. Therefore, rejecting H_0 at significance level 0.05 is equivalent to the inclusion of the two-sided 90% confidence interval for $\exp(\beta_1)$ in the margin (0.8, 1.25).

It is well known that AUC and C_{max} are strongly correlated because they are two inter-related measures that represent the extent and speed of the exposure to a drug, respectively. The correlation between AUC and C_{max} could reduce the intra-individual variability of each measure. Traditionally, however, the 90% confidence intervals of GMR between a test product and the reference product have been estimated separately for AUC and C_{max} assuming these two endpoints are independent. This could result in wider 90% confidence intervals, thereby increasing the probability that a test product fails to show bioequivalence to its reference product when, in fact, it is bioequivalent.

'Highly variable drugs' have been defined as those drugs for which the within-subject variability equals or exceeds 30% of the C_{max} or AUC ($i.e.$ $CV_w \geq 0.3$). The bioequivalence of highly variable drugs is a problem because the high variability brings a wide 90% confidence interval for GMR which does not be included in the bioequivalence margin (0.8, 1.25). To deal with the problems posed by highly variable drugs, the regulatory agencies permit to broaden the bioequivalence acceptance limits for C_{max}, e.g the 90% CI for the GMR of C_{max} values might be required to fit within acceptance limits of (0.75, 1.33) or even (0.70,1.42).

Multivariate hierarchical generalized linear model (multivariate HGLM) is a hierarhcical likelihood (h-likelihood) approach to simultaneously deal with multiple endpoints. For example, Molas *et al.* (2013) developed a joint model for three correlated outcomes in rheumatoid arthritis using multivariate Gaussian random effects. Double hierarchical generalized linear model (double HGLM) is a class of models proposed by Lee and Nelder (2006) in which random effects can be specified for both the mean and dispersion. This class enable models with heavy-tailed distributions to be explored,

providing robust estimation agianst outliers. Multivarite double hierarchical generalized linear models (multivariate double HGLMs) can be applied to the study of bioequivalence where it is of interest to make common conclusions (bioequivalence) for both AUC and C_{max}, even in the case of highly variable drugs.

2.2 Motivating examples

2.2.1 Example 1: Tramadol data

The first example is the comparative pharmacokinetic study between extended-release (ER) and conventional immediate-release (IR) of tramadol/acetaminophen fixed-dose combination, which is an open-label, multiple dose, randomized, 2-sequence 2-period crossover study conducted in 12 healthy male volunteers. All subjects received both formulations: either IR tablet for 4 days followed by ER tablet for 4 days, or vice versa. A 5-day washout period separated the two treatments. For the comparison of the pharmacokinetic characteristics between ER and IR treatments, the C_{max} and AUC_{0-12} at steady state (on day 4) were analyzed by the linear mixed model (2.1) with the period, sequence, and treatment as fixed effects and subject as random effects.

The GMR of IR to ER formulation for AUC_{0-12} at steady state was 0.947 with 90% CI (0.910, 0.986). However, those for C_{max} was 0.866 with 90% CI (0.797, 0.941) which was slightly out of the bioequivalence limit (0.8, 1.25). Figure 2.2 shows that there exists strong positive correlations between subject-specific random effects for log(C_{max}) and those for log(AUC) in tramadol data (correlation coeffecient r=0.966). Therefore, C_{max} and AUC should be analyzed simultaneously using multivariate HGLM in which the correlation between outcomes is specified.

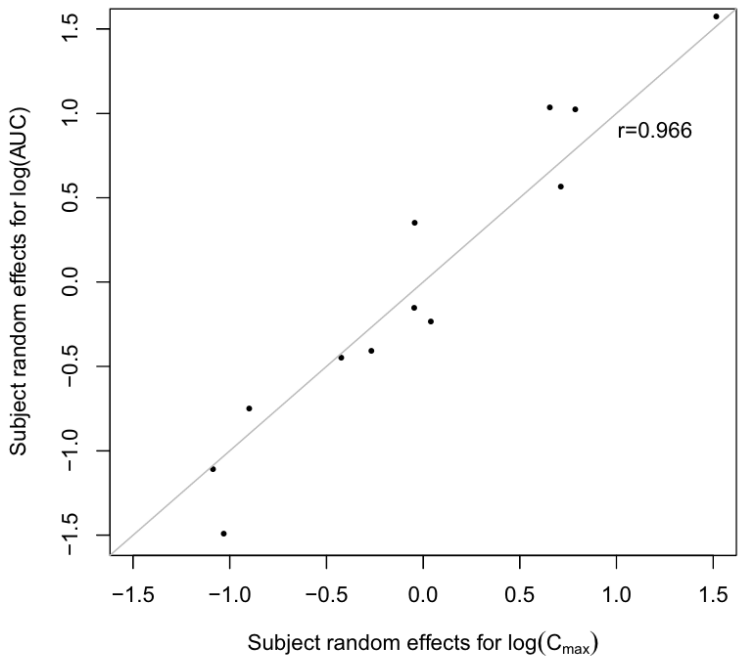

Figure 2.2: Scatter plots of seperately predicted subject random effects for $\log(C_{max})$ and log(AUC) using the linear mixed model (2.1) and their correlation coefficient r for tramadol data of example 1.

2.2.2 Example 2: Fimasartan data

The second example is the drug-drug interaction study between fimasartan and amlodipine, which is an a open-label, multiple-dose, 1-sequnece 2-period study conducted in 19 healthy male volunteers. All subjects were administered fimasartan alone for 7 days in period I, and after a 5-day washout period, they received amlodipine coadministered with fimasartan for 7 days in period II. To investigate the effect of the coadministration of fimasartan and amlodipine on the steady-state pharmacokinetics, the C_{max} and AUC_{0-24} at steady state on day 7 were analyzed by linear mixed models with the treatment as fixed effects and subject as random effects.

The GMR (coadministration of fimasartan with amlodipine to fimasartan alone) and its 90% CI for C_{max} were 1.096 (90% CI: 0.761, 1.579), and those for AUC_{0-24} were 1.163 (90% CI: 1.009, 1.341). This indicates that fimasartan belongs to highly variable drugs with large within-subject variabilities for C_{max} ($CV_w = 0.77$) and moderate within-subject variance for AUC ($CV_w = 0.27$). Similar to the tramadol data of Example 1, the fimasartan data have strong positive correlations between subject-specific random effects for log(C_{max}) and those for log(AUC) in Figure 2.3. Therefore, we can consider a dispersion model for the large within-subject variance of C_{max} as well as correlations between C_{max} and AUC. Multivariate double hierarchical generalized linear models can be applied to the fimasartan data.

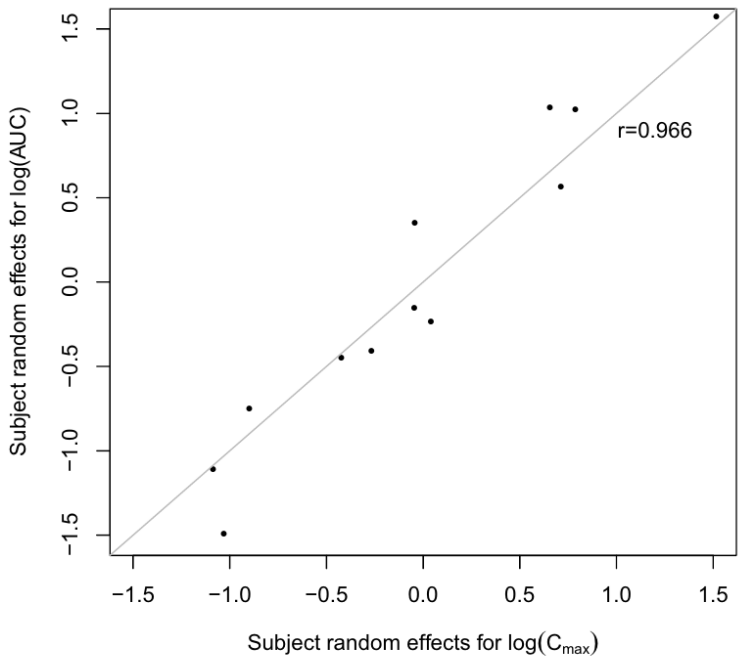

Figure 2.3: Scatter plots of seperately predicted subject random effects for $\log(C_{max})$ and log(AUC) using the linear mixed model (2.1) and their correlation coefficients r for fimasartan data of example 2

2.3 Multivariate double hierarchical generalized linear models and h-likelihood theory

2.3.1 Multivariate double hierarchical generalized linear models

Let y_{ij} be the jth outcome from cluster i ($i = 1, \cdots, m; j = 1, \cdots, q_i$). Within a cluster, y_{ij}'s are dependent, but conditional on the cluster-specific random effects v_i, the response y_{ij}'s are independent. Then we can consider a linear mixed model given by,

$$y_{ij} = X_{ij}^T \beta + Z_{ij}^T v_i + e_{ij} \tag{2.2}$$
$$v_i \sim N(0, D),\ e_{ij} \sim N(0, \phi_{ij}),$$

where X_{ij} and Z_{ij} are the covariate vectors for the fixed effects β and the random effects v_i of cluster i. When we fit this model (2.2) for two co-primary endponts $\log(\text{AUC})$ and $\log(\text{C}_{\text{max}})$ seperately, then two independent 90% confidence intervals for GMR of the test product to the reference can be obtained for the bioequivalence testing. This is the conventional method to test the bioequivalence.

When the within-cluster variance is high such as highly variable drugs, we can consider a dispersion model for the within-cluster variance ϕ_{ij} containing random effects with log-link:

$$\log(\phi_{ij}) = G_{ij}^T \gamma + F_{ij}^T b_i \tag{2.3}$$
$$b_i \sim N(0, \alpha I),$$

where G_{ij} and F_{ij} are the covariate vectors for the fixed effects γ and the random effects b_i of cluster i. This model (2.3) becomes a double hierarchical genelized linear model (Lee and Nelder, 2006). Similarly, we can also consider a dispersion model for λ, the dispersion parameter vector for $\text{var}(v_i) = D = D(\lambda)$. However, between-cluster variations are not of interest for clinical pharmacology data, we consider the model having random effects in residual variances ϕ_{ij} only.

If we have two or more co-primary endpoints which are correlated such as C_{max} and AUC in clinical pharmacology data, multivariate double hierarchical generalized linear models can be applied (Lee *et al.*, 2017). Let y_{1ij} and y_{2ij} be jth two outcomes from cluster i ($i = 1, \cdots, m; j = 1, \cdots, q_i$), and v_{1i} and v_{2i} be the cluster-specific random effects which are correlated. It is assumed that y_{1ij} and y_{2ij} are conditionally independent given v_{1i} and v_{2i}. Then the following bivariate double HGLM can be proposed:

$$
\begin{aligned}
y_{1ij}\big|v_{1i}, b_{1i} &\sim N(\mu_{1ij}, \phi_{1ij}), \\
\mu_{1ij} &= X_{1ij}^T\beta + Z_{1ij}^T v_{1i}, \\
\log(\phi_{1ij}) &= G_{1ij}^T\gamma_1 + F_{1ij}^T b_{1i}, \\
b_{1i} &\sim N(0, \alpha_1 I),
\end{aligned}
\tag{2.4}
$$

$$
\begin{aligned}
y_{2ij}\big|v_{2i}, b_{2i} &\sim N(\mu_{2ij}, \phi_{2ij}), \\
\mu_{2ij} &= X_{2ij}^T\beta + Z_{2ij}^T v_{2i}, \\
\log(\phi_{2ij}) &= G_{2ij}^T\gamma_2 + F_{2ij}^T b_{2i}, \\
b_{2i} &\sim N(0, \alpha_2 I),
\end{aligned}
$$

$$
\begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix} \sim \mathbf{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}
$$

where X_{kij} and Z_{kij} are the covariate vectors for the fixed effects β_k and the random effects v_{ki} of cluster i for mean models, and G_{kij} and F_{kij} are the covariate vectors for the fixed effects γ_k and the cluster-specific random effect b_{ki} for dispersion models ($k = 1, 2$). Within a cluster i, the random effects v_{ki} for mean models and the random effects b_{ki} for dispersion models are independent. Correlations between y_{1ij} and y_{2ij} within a cluster i are explained through the correlations between cluster-specific random effects v_{1i} and v_{2i}. The variance-covariance matrix Σ of v_{1i} and v_{2i} depends on the dispersion parameter vector λ which consist of correlation coeffecients

and diagonal variance components (i.e., $\Sigma = \Sigma(\lambda)$). For example, let Σ be the 2×2 variance-covariance matrix such that $\Sigma = \begin{pmatrix} \lambda_1 & \rho\sqrt{\lambda_1\lambda_2} \\ \rho\sqrt{\lambda_1\lambda_2} & \lambda_2 \end{pmatrix}$ with the dispersion parameter vector $\lambda = (\rho, \lambda_1, \lambda_2)$. If the correlation coefficient $\rho = 0$, then two outcomes y_{1ij} and y_{2ij} from the same cluster are independent.

2.3.2 H-likelihood estimation procedure

Consider the multivariate double hierarchical generalized linear model given by (2.4). To set the notation, let y_k denote the outcome vector of kth endpoint stacked over subject, similarly X_k and Z_k denote the design matrices of fixed effects β_k and random effects v_k $(k = 1, 2)$.

The h-likelihood of the model (2.4) is

$$\begin{aligned} h &= l_1(\beta_1, \gamma_1, \beta_2, \gamma_2; y_1, y_2|v_1, b_1, v_2, b_2) + l_2(\lambda; v_1, v_2) + l_3(\alpha_1, \alpha_2; b_1, b_2) \\ &= -\frac{1}{2}\log|2\pi\Phi_1| - \frac{1}{2}\left(y_1 - X_1\beta_1 - Z_1v_1\right)^T {\Phi_1}^{-1}\left(y_1 - X_1\beta_1 - Z_1v_1\right) \\ &\quad -\frac{1}{2}\log|2\pi\Phi_2| - \frac{1}{2}\left(y_2 - X_2\beta_2 - Z_2v_2\right)^T {\Phi_2}^{-1}\left(y_2 - X_2\beta_2 - Z_2v_2\right) \\ &\quad -\frac{1}{2}\log|2\pi\Sigma_R| - \frac{1}{2}v^T\Sigma_R^{-1}v \\ &\quad -\frac{m}{2}\log(2\pi\alpha_1) - \frac{1}{2\alpha_1}{b_1}^T b_1 \\ &\quad -\frac{m}{2}\log(2\pi\alpha_2) - \frac{1}{2\alpha_2}{b_2}^T b_2 \end{aligned}$$

where Φ_k is a diagonal matrix of $\phi_{kij} = \exp(G_{kij}^T\gamma_k + F_{kij}^T b_{ki})$ $(k = 1, 2)$ and Σ_R is the variance-covariance matrix of random effects vector $v = (v_1^T, v_2^T)^T$ with dispesion parameter vector λ such that

$$\Sigma_R = \Sigma_R(\lambda) = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \otimes \mathbf{I}_m.$$

For a likelihood l with nuisance effects δ, Lee and Nelder (2001) considered a class of adjusted profile likelihood $p_\delta(l)$, defined by

$$p_\delta(l) = \left[l - \frac{1}{2} \log \det \left\{ \frac{(D(l,\delta)}{2\pi} \right\} \right]_{\delta=\hat{\delta}},$$

where $D(l,\delta) = -\partial^2 l / \partial\delta\partial\delta^T$ and $\hat{\delta}$ solves $\partial l / \partial\delta = 0$. This class is obtained from eliminating nuissance effects δ which may stand for fixed or random or both. Suppose δ is a vector representing random effects, then $p_\delta(l)$ can be shown to be Laplace approximation to the marginal log-likelihood $L_v = \int \exp(h) dv$ by integrating out the nuisance random effects. Lee and Nelder (2001) showed that this form used in Laplace approximation in identical to that of the adjusted profile likelihood to eliminate fixed parameters by conditioning on maximum likelihood estimator $\hat{\delta}$ under the parameter orthogonality (Cox and Reid, 1987). This enables the extension of restreicted maximum likelihood estimators. For example, $p_{\beta,v}(h)$ is an extended restricted likelihood eliminating nuisance random effect v by intergration, while eliminating nuisance fixed effects β by conditioning on $\hat{\beta}$.

We use h-likelihood for inference about v, the marginal likelihood L_v for β, the restricted likelihood $p_\beta(L_v)$ for the dispersion components γ, b, λ, and the extended restricted likelihood $p_{\beta,\gamma}(L_{v,b})$ for α where $L_{v,b} = \int \exp(h) dv db$. In general, $p_{\beta,v}(h)$ is approximately $p_\beta\{p_v(h)\}$ and therefore $p_\beta(L_v)$. When L_v and $L_{v,b}$ are numerically difficult to obtain, $p_v(h)$, $p_{\beta,v}(h)$, and $p_{\beta,v,\gamma,b}(h)$ are used as approximations to L_v, $p_\beta(L_v)$, and $p_{\beta,\gamma}(L_{v,b})$, repectively. In the multivariate double hierarchical generalized linear model (2.4), the adjusted term $D(h,v)$ in $p_v(h)$ does not depend on v, then both v and β are estimated using h as shown in Table 2.1.

Table 2.1: Criteria for effects in multivariate double HGLMs (2.4)

Criterion	Estimated	Eliminated	Approximation
h	β, v		
$p_\beta(L_v)$	γ, b, λ	β, v	$p_{\beta,v}(h)$
$p_{\beta,\gamma}(L_{v,b})$	α	β, v, γ, b	$p_{\beta,v,\gamma,b}(h)$

H-likelihood estimation procedure for the model (2.4) is the following:

1. Set dispersion components at some initial values.
2. Given dispersion components, estimate $\hat{v} = \begin{pmatrix} \hat{v}_1 \\ \hat{v}_2 \end{pmatrix}$, and $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$ maximazing h.
3. Given $\hat{\beta}$ and $\hat{v}$ from 2, estimate new $\hat{\gamma} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix}$ and $\hat{b} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}$ maximazing $p_{\beta,v}(h)$, and then estimate new $\hat{\alpha}$ maximizing $p_{\beta,v,\gamma,b}(h)$.
4. Estimate new $\hat{\lambda}$ maximazing $p_{\beta,v}(h)$ of 3.
5. Iterate steps $2-4$ until convergence.

For details of the estimation procedures, see Lee *et al.* (2017).

2.4 Application for clinical pharmacology data

2.4.1 Tramadol data

For the tramadol data of the motivating example 1, we can consider the following bivariate hierarchical generalized linear model (bivariate HGLM): Let y_{1ij} and y_{2ij} be logarithms of C_{max} and AUC obtained from the subject i in the jth period with 2×2 crossover design, and $(v_{1i}, v_{2i})^T$ be the subject-specific random effects ($i = 1, \cdots, 12; j = 1, 2$). It is assumed that y_{1ij} and y_{2ij} are conditionally independent given $(v_{1i}, v_{2i})^T$.

$$
\begin{aligned}
&y_{1ij}\big|v_{1i} \sim N(\mu_{1ij}, \phi_1),\\
&\mu_{1ij} = \beta_{10} + \beta_{11}\, trt_{ij} + \beta_{12}\, prd_j + \beta_{13}\, seq_i + v_{1i},
\end{aligned}
$$

and

$$
\begin{aligned}
&y_{2ij}\big|v_{2i} \sim N(\mu_{2ij}, \phi_2),\\
&\mu_{2ij} = \beta_{20} + \beta_{21}\, trt_{ij} + \beta_{22}\, prd_j + \beta_{23}\, seq_i + v_{2i},
\end{aligned}
$$

and the subject random effects follow a multivariate normal distribution

$$
\begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix} \sim \mathbf{N}\left[0, \begin{pmatrix} \lambda_1 & \rho\sqrt{\lambda_1\lambda_2} \\ \rho\sqrt{\lambda_1\lambda_2} & \lambda_2 \end{pmatrix}\right],
$$

where $trt_{ij} = 1$ for the ER formulation and $trt_{ij} = 0$ for the IR formulation, prd_j and seq_i are covariates indicating period j and sequence of the subject i, and the correlation coefficient ρ of subject random effects satisfies $-1 \leq \rho \leq 1$. If $\rho = 0$, then it is equivalent to conventional independent two linear mixed models.

We performed the analysis to evaluate the bioequivalence of two formulations ER and IR. The summaries of the analysis are presented in Table 2.2. Based on the restricted likelihood $p_{\beta,v}(h)$, the likelihood ratio, $LR = -2\{p_{\beta,v}(h)\big|_{H_0} - p_{\beta,v}(h)\big|_{H_1}\}$ is formed to test the correlation coeffecient ρ is statistically significant or not. The

likelihood ratio test for $H_0 : \rho = 0$ rejects the null hypothesis because the deviance difference between independent models and the bivariate HGLM is 12.6 $(= -31.3 + 43.9) > \chi^2_{0.05}(1) = 3.84$, thus likelihood ratio test selects the bivariate HGLM. The standard error for $\hat{\beta}_{11}$ of the biavariate HGLM is 0.040 which is smaller than 0.046, that of the independent models, and this leads to the narrower 90% confidence interval of GMR for C_{max}. As a result of the bivariate HGLM, both 90% confidence interval of GMR for C_{max} and AUC are included into the bioequivalence margin (0.8, 1.25), then the ER formulation is equivalent to the IR formulation.

2.4.2 Fimasartan data

From the result of conventional linear mixed models for the fimasartan data of example 2, fimasartan belongs to the highly variable drug with large within-subject coefficient of variation $CV_w = 0.77$ for C_{max}, while $CV_w = 0.27$ for AUC. Therefore, to account the heterogeneity of within-subject variance (the residual error variance) we can consider the following bivariate double hierarchical generalized linear model (bivariate double HGLM) with a dispersion model for C_{max}. Let y_{1ij} and y_{2ij} be logarithms of C_{max} and AUC obtained from the subject i in the jth treatment, and v_{1i}, v_{2i} be the subject-specific random effects $(i = 1, \cdots, 19; j = 1, 2)$. It is assumed that y_{1ij} and y_{2ij} are conditionally independent given v_{1i}, v_{2i}.

$$
\begin{aligned}
y_{1ij} \big| v_{1i}, b_{1i} &\sim N(\mu_{1ij}, \phi_{1ij}), \\
\mu_{1ij} &= \beta_{10} + \beta_{11}\, trt_j + v_{1i}, \\
\log(\phi_{1ij}) &= \gamma_{10} + b_{1i}, \\
b_i &\sim N(0, \alpha_1),
\end{aligned}
$$

and

$$y_{2ij}\big|v_{2i} \sim N(\mu_{2ij}, \phi_2),$$

$$\mu_{2ij} = \beta_{20} + \beta_{21}\, trt_j + v_{2i},$$

and the subject random effects follow a multivariate normal distribution

$$\begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix} \sim \mathbf{N}\left[0, \begin{pmatrix} \lambda_1 & \rho\sqrt{\lambda_1\lambda_2} \\ \rho\sqrt{\lambda_1\lambda_2} & \lambda_2 \end{pmatrix}\right],$$

where $trt_{ij} = 1$ for the coadministration of fimasartan with amlodipine and $trt_{ij} = 0$ for the fimasartan alone and $-1 \leq \rho \leq 1$. We consider four models as follows:

(i) M1: bivariate double HGLM with a dispersion model

$$\log(\phi_{1ij}) = \gamma_{10} + b_{1i}, \;\; b_i \sim N(0, \alpha_1).$$

(ii) M2: independent double HGLMs with a dispersion model

$$\log(\phi_{1ij}) = \gamma_{10} + b_{1i}, \;\; b_i \sim N(0, \alpha_1),$$

where $\rho = 0$.

(iii) M3: bivariate HGLM with $\phi_{1ij} = \phi_1$.

(iv) M4: independent linear mixed models where $\rho = 0$ and $\phi_{1ij} = \phi_1$.

M1 is the full model and the others are various simplications of it by assuming null components, i.e., M2 ($\rho = 0$), M3 ($\alpha_1 = 0$), M4 ($\rho = 0, \alpha_1 = 0$). Because M2 and M3 are non-nested models, we use conditional Akaike information criterion (cAIC) to select the best fitting model among four models. M1: bivariate double HGLM is selected, which has the smallest cAIC=103.99 among four models. The details of cAIC for double HGLMs are given in chapter 3.

The summaries of results from four models are in Table 2.3. M1 : bivariate double HGLM gives the smallest standard errors for treatment efffects $\hat{\beta}_{11}$ and $\hat{\beta}_{21}$, which results in the narrowest 90% confidence intervals.

Table 2.2: Summaries of analysis for the tramadol data.

	bivariate HGLM		Independent models ($\rho = 0$)	
	Estimate	S.E.	Estimate	S.E.
C_{max} β_{11}	-0.144	0.040	-0.144	0.046
AUC β_{21}	-0.054	0.022	-0.054	0.022
ρ	0.884		0	
$-2\, p_{\beta,v}(h)$	-43.9		-31.3	
	GMR	(90% CI)	GMR	(90% CI)
C_{max}	0.866	(0.805, 0.931)	0.866	(0.797, 0.941)
AUC	0.947	(0.910, 0.986)	0.947	(0.910, 0.986)

Table 2.3: Summaries of analysis for the fimasartan data.

		M1: bivariate double HGLM		M2: independent double HGLMs	
		Estimate	S.E.	Estimate	S.E.
C_{max}	β_{11}	0.088	0.173	0.086	0.203
AUC	β_{21}	0.151	0.082	0.151	0.086
	ρ	0.831		0	
		GMR	(90% CI)	GMR	(90% CI)
C_{max}		1.091	(0.821, 1.451)	1.090	(0.781, 1.521)
AUC		1.163	(1.016, 1.332)	1.163	(1.009, 1.341)
		M3: bivariate HGLM		M4: independent LMMs	
		Estimate	S.E.	Estimate	S.E.
C_{max}	β_{11}	0.092	0.191	0.092	0.221
AUC	β_{21}	0.151	0.082	0.151	0.086
	ρ	0.832		0	
		GMR	(90% CI)	GMR	(90% CI)
C_{max}		1.096	(0.801, 1.501)	1.096	(0.762, 1.578)
AUC		1.163	(1.016, 1.332)	1.163	(1.009, 1.341)

Chapter 3

Conditional Akaike information for double hierarchical generalized linear models

3.1 Literature review

Akaike (1973) based his information on the Kullback-Leibler distant, given by

$$I(f, g_\theta) = E_f \log f(y) - E_f \log g_\theta(y)$$

where f is true density generating data y and $g_\theta = g(\cdot|\theta)$ is a parametrized model for $\theta \in \Theta$. Smaller values of $I(f, g_\theta)$ correspond to a better approximation of f by g_θ, and the miminum is obtained for some θ_0, the pseudotrue parameter. If the true f belongs to the parametrized class of models $\mathbb{G} = \{g_\theta; \theta \in \Theta\}$, then $f = g_{\theta_0}$ and $I(f, g_{\theta_0}) = 0$. In practice θ_0 is estimated as $\hat{\theta}$ from the data y, and $I(f, g_{\theta_0})$ is approximated by $I(f, g_{\hat{\theta}})$, where $\hat{\theta} = \hat{\theta}(y)$ is usually the maximum likelihood estimator based on the data y. The quality of the approximation of the true f by the class of models $\mathbb{G}$ is assessed by the quantity

$$E_f I(f, g_{\hat{\theta}}) = E_{f(y^*)} \log f(y^*) - E_{f(y)} E_{f(y^*)} \log g\{y^*|\hat{\theta}(y)\},$$

where y^* is generated from f and independent to y. When we compare different classes of models, the constant term $E_{f(y^*)} \log f(y^*)$ can be ignored, and the relative fit of

competing models can be assessed using the Akaike information (AI)

$$\text{AI} = -2E_{f(y)}E_{f(y^*)} \log g\{y^*|\hat{\theta}(y)\}.$$

The Akaike information criterion (AIC) is an estimator of the AI, given by

$$\text{AIC} = -2 \log g\{y|\hat{\theta}(y)\} + 2p$$

where p is the number of free parameters in model class $\mathbb{G}$ (i.e., degree of freedom). When $\hat{\theta}(y)$ is the maximum likelihood estimator (MLE) and f is included in the parametrized class of models $\mathbb{G}$, then AI $= E(\text{AIC}) + o(1)$ (Burnham amd Anderson, 2002).

3.1.1 Effective degree of freedom

Lee and Nelder (1996) proposed the scaled deviance for the hierarchical generalized linear model (HGLM) in which random effects were specified, based on the conditional log-likelihood $l_1(\mu; y|v) = \log g(y|v; \mu)$ for y given random effects v,

$$D(\hat{\mu}) = -2l_1(\hat{\mu}; y|v) + 2l_1\{y; y|v\},$$

with the estimated degree of freedom $n - p_D$, where the conditional mean $\mu = E(y|v)$ depends on fixed effects β and random effects v in the model. The quantity

$$p_D = tr(H_1^{-1}H_1^*) \tag{3.1}$$

is computed using Hessian matrices

$$H_1 = -\frac{\partial^2 h}{\partial(\beta, v)\partial(\beta, v)^T} \text{ and } H_1^* = -\frac{\partial^2 l_1}{\partial(\beta, v)\partial(\beta, v)^T},$$

for the h-likelihood and the conditional log-likelihood with respect to the parameters β and v. Note that p_D depens on β and v, and so it is evaluated using their estimates $\hat{\beta}$ and $\hat{v}$.

When the model does not contain random effects, p_D becomes the rank of the model design matrix in the corresponding generalized linear model (GLM). Therefore, the quantity $p_D = tr(H_1^{-1}H_1^*)$ is an 'effective degrees of freedom' adjusted to estimate both fixed and random effects (Ha *et al.*, 2007). This coinsided with the generalized degree of freedom (GDF) proposed by Ye (1998). Speigelhalter *et al.* (2002) viewed p_D as a Bayesian measure of model complexity. In particular, $p_D = tr(P_1)$ for linear mixed models with known variances, where P_1 is the 'Hat-matrix' satisfying $\hat{y} = X\hat{\beta} + Z\hat{v} = P_1 y$ with MLE $\hat{\beta}$ and empirical Bayes estimator (EBE) $\hat{v} = E(v|y)$ (Hodge and Sargent, 2001).

3.1.2 Conditional Akaike information criterion

Assume that the true conditional distribution of data y is $f_1(y|u)$, and that u is the true random effects vector with distribution $f_2(u)$. The prediction data set y^* is independent to y conditional on u, and from the same distribution $f_1(\cdot|u)$. In other words, y and y^* share the same random effects u, but differ in their error terms. The conditional Akaike information (cAI) is defined to be

$$\begin{aligned} \text{cAI} &= -2E_{f(y,u)}E_{f(y^*|u)} \log g\{y^*|\hat{\beta}(y), \hat{v}(y)\} \\ &= \int -2 \log g\{y^*|\hat{\beta}(y), \hat{v}(y)\} f_1(y^*|u) f(y, u,) dy^* dy du, \end{aligned}$$

where $f(y, u) = f_1(y|u)f_2(u)$ is the true joint distribution of y and u , $\hat{\beta}(y)$ and $\hat{v}(y)$ are estimators of fixed effects β and random effects v based on data y (Vaida and Blanchard, 2005).

The conditional Akaike information criterion (cAIC) is an estimator of the cAI, given by

$$\text{cAIC} = -2 \log g\{y|\hat{\beta}(y), \hat{v}(y)\} + 2p,$$

where the bias correction term p is realted to the number of parameters in the model (i.e., effective degree of freedom). Because this bias correction term p determines the

bias of cAIC, its unbiased esimator $\hat{p}$ have been developed for various models such as linear mixed models and generalized linear mixed models under many different conditions.

In the linear mixed model with known variances, Vaida and Blanchard (2005) showed that conditional AIC with $\hat{p}_c = tr(P_1)$ is asymptotically unbiased, where P_1 is the 'Hat-matrix' satisfying $\hat{y} = X\hat{\beta} + Z\hat{v} = P_1 y$ with MLE $\hat{\beta}$ and EBE $\hat{v} = E(v|y)$. In generalized linear mixed models with known variances, using h-likelihood estimators $\hat{\beta}$, $\hat{v}$ and the bias correction estimatior $\hat{p}_c = tr(H_1^{-1}H_1^*)$, conditional AIC is asymptotically unbiased (Donohue *et al.*, 2011; Yu and Yau, 2012). This bias correction estimator $\hat{p}_c$ accurately corresponds to the effective degree of freedom p_D in (3.1).

In the case of unknown variances, let ϕ and λ be the dispersion parameters for the variance components var$(y|v)$ and var(v), respectively. In practice, Vaida and Blanchard (2005) suggested using plug-in estimators $\hat{\phi}$ and $\hat{\lambda}$ for unknown variances, and adding 1 to the effective degree of freedom such that $\hat{p} = tr(P_1) + 1$ for linear mixed model. Because the conditional log-likelihood, $\log g(y|v)$, does not depend on λ, they only considered uncertainty in estimation of dispersion parameter ϕ by adding 1 to the effective degree of freedom, ignoring those of $\hat{\lambda}$. However, Greven and Kneib (2010) showed that this plug-in method induced a very specific bias of cAIC: corresponding cAIC tends to select a random effect into the model. A similar phenomenon was also found in Ha *et al.* (2007).

Yu and Yau (2012) proposed the corrected version of cAIC for generalized linear mixed models using a different asymptotic unbiased estimator for the bias correction term p, given by

$$\hat{p}_{ml} = tr\left\{\left(H_{\theta\theta} - H_{\theta\lambda}H_{\lambda\lambda}^{-1}H_{\lambda\theta}\right)^{-1} H_{\theta\theta}^*\right\}, \tag{3.2}$$

where $H_{\theta\theta}$ and $H_{\theta\theta}^*$ are Hessian matrices of h-likelihood and conditional log-likelihood with respect to the parameter $\theta = (\beta^T, v^T)^T$. The matrices $H_{\theta\lambda}$ and $H_{\lambda\lambda}$ are the corresponding Hessian matries of the adjusted likelihood $h_a = h - 0.5\log\det\left(-\partial^2 h/\partial v \partial v^T\right)$

with respect to the parameters θ and λ. This bias correction estimator considered the uncertainty in estimation of random effects variance λ. They showed that the estimator $\hat{p}_{ml}$ under known dispersion parameter ϕ is asymptotically equivalent to $\Phi_0(y)$ in Liang *et al.* (2008). However, they did not cover the case of unknown dispersion ϕ. When we consider joint models for both mean and dispersion such as joint GLMs and double HGLMs, the bias correction estimator $\hat{p}$ of cAIC should take the uncertainty in estimation of both dispersion parameters ϕ and λ into consideration.

3.1.3 Other Akaike information criteria

In addition to the conditional Akaike information criterion based on the conditional log-likelihood, we may use two other AIC (Ha *et al.*, 2007) based on marginal log-likelihood m and restricted log-likelihood r, defined by

$$\begin{aligned} \text{mAIC} &= -2\log m + 2df_m \approx -2\log p_v(h) + 2df_m, \\ \text{rAIC} &= -2\log r + 2df_r \approx -2\log p_{\beta,v}(h) + 2df_r, \end{aligned}$$

where df_m is the number of fixed parameters and df_r is the number of dispersion parameters. For example, in linear mixed models cAIC focuses on the prediction at the cluster level $\hat{y} = \hat{\mu} = X\hat{\beta} + Z\hat{v}$, where $\mu = E(y|v)$, whereas mAIC on the marginal prediction $\hat{y} = X\hat{\beta}$. When we compare models with different fixed parameters, mAIC can be used, whereas rAIC can be used for dispersion parameter model selection. However, these two AIC based on marginal and restricted log-likelihood can not be used to select models involving both fixed and random effects.

3.2 Conditional Akaike information for double hierarchical generalized linear model

3.2.1 h-likelihood inference

Recall the double hierarchical generalized linear model (double HGLM) with a dispersion model containing cluster-specific random effects b_i with log-link

$$\begin{aligned} y_{ij} &= X_{ij}^T\beta + Z_{ij}^T v_i + e_{ij}, \\ v_i &\sim N(0,\lambda_i),\ e_{ij} \sim N(0,\phi_{ij}), \\ \log(\phi_{ij}) &= G_{ij}^T\gamma + F_{ij}^T b_i, \\ b_i &\sim N(0,\alpha I), \end{aligned} \tag{3.3}$$

where X_{ij} and Z_{ij} are the covariate vectors for the fixed effects β and the cluster-specific random effects v_i from the mean model, and G_{ij} and F_{ij} are the covariate vectors for the fixed effects γ and the cluster-specific random effects b_i from the dispersion model ($i = 1, \cdots, m$; $j = 1, \cdots, q$). Note that $n = mq$.

Let the stack function stacks matrices on top of each other. The distribution of random effects $v = \text{stack}(v_1, v_2, \cdots, v_m)$ from the mean model is multivariate normal with $E(v) = 0$ and $\text{var}(v) = \Lambda = \text{diag}(\lambda_i)$. Let X_i and Z_i be the matrices with rows X_{ij}^T and Z_{ij}^T, and let $X = \text{stack}(X_1, X_2, \cdots, X_m)$ and $Z = \text{diag}(Z_1, Z_2, \cdots, Z_m)$ be the $n \times p$ and $n \times m$ model matrices. For random efffects $b = \text{stack}(b_1, b_2, \cdots, b_m)$ of the dispersion model, $E(b) = 0$ and $\text{var}(b) = \alpha I$. Similarly, let G_i and F_i be the matrices with rows G_{ij}^T and F_{ij}^T, and let $G = \text{stack}(G_1, G_2, \cdots, G_m)$ and $F = \text{diag}(F_1, F_2, \cdots, F_m)$ be the $n \times r$ and $n \times m$ model matrices. We can also consider a dispersion model for $\lambda = (\lambda_1, \cdots, \lambda_m)^T$. For simplicity of argument, we consider the model having random effects in dispersion ϕ_{ij} only.

The h-likelihood of the model (3.3) is

$$\begin{aligned} h &= l_1(\beta,\gamma;y|v,b) + l_2(\lambda;v) + l_3(\alpha;b) \\ &= -\frac{1}{2}\log|2\pi\Phi| - \frac{1}{2}(y-X\beta-Zv)^T\Phi^{-1}(y-X\beta-Zv) \\ &\quad -\frac{1}{2}\log|2\pi\Lambda| - \frac{1}{2}v^T\Lambda^{-1}v \\ &\quad -\frac{m}{2}\log|2\pi\alpha| - \frac{1}{2\alpha}b^Tb \end{aligned}$$

where Φ is a diagonal matrix of $\phi_{ij} = \exp(G_{ij}^T\gamma + F_{ij}^T b_i)$.

The model (3.3) can be viewed as an augmented linear model as

$$\begin{aligned} y_a &= T\theta + e_a, \\ e_a &\sim N(0,\Sigma_a), \end{aligned} \tag{3.4}$$

where the augmented response $y_a = \text{stack}(y,\psi_M)$ contains y and pseudo-response $\psi_M = 0$ such that $E(\psi_M) = v$ and $\text{var}(\psi_M) = \Lambda$. Therefore, the variance matrix of the augmented error e_a becomes block-diagonal such that $\Sigma_a = \begin{pmatrix} \Phi & 0 \\ 0 & \Lambda \end{pmatrix}$. The design matrix $T = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix}$ and the parameter vector $\theta = \text{stack}(\beta, v)$ includes both fixed and random effects. The cluster-specific random effects can be correlated, for example, $\text{var}(v) = \lambda A$ for a given correlation matrix A. However, the model can be transformed into the independent random effects by recomputing Z as $ZA^{\frac{1}{2}}$ where $A^{\frac{1}{2}}$ is some square-root transformation of A (e.g., Cholesky decomposition). Thereby, the above augmented model having independent random effects can be applied. See Chapter 8 of Lee et al. (2017) for further details.

Estimating equations for $\hat{\beta}$ and $\hat{v}$ of model (3.3) are

$$\begin{pmatrix} X^T\Phi^{-1}X & X^T\Phi^{-1}Z \\ Z^T\Phi^{-1}X & Z^T\Phi^{-1}Z + \Lambda^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{v} \end{pmatrix} = \begin{pmatrix} X^T\Phi^{-1}y \\ Z^T\Phi^{-1}y \end{pmatrix},$$

which is equivaletnt to $T^T\Sigma_a^{-1}T\hat{\theta} = T^T\Sigma_a^{-1}y_a$ of the augmented model (3.4). The h-liklihood estimator $\hat{\beta}$ and $\hat{v}$ satisfying the above estimating equation are the maximum

likelihood estimator for β and the empirical Bayes estimator for v, respectively (Lee and Nelder, 1996).

For estimating variance components γ, b, and λ, we use the adjusted profile likelihood

$$p_{\beta,v}(h) = \left[h - \frac{1}{2} \log\det\left(\frac{D\{h,(\beta,v)\}}{2\pi} \right) \right]_{(\beta,v)=(\hat{\beta},\hat{v})},$$

where $D\{h,(\beta,v)\} = -\partial^2 h \big/ \partial(\beta,v)\partial(\beta,v)^T$ and $(\hat{\beta},\hat{v})$ solves $\partial h/\partial(\beta,v) = 0$. This adjusted profile likelihood $p_{\beta,v}(h)$ approximates the restricted likelihood $p_\beta(L_v)$ where $L_v = \int \exp(h) dv$.

Let ϕ_k and p_k be the k-th diagonal elements of Φ and $P = T(T^T\Sigma_a^{-1}T)^{-1}T^T\Sigma_a^{-1}$, and d_k be the square of the k-th element of $\hat{e} = y - X\hat{\beta} - Z\hat{v}$. Consider a parametrization of $\phi = (\phi_1, \phi_2, \cdots, \phi_n)^T$ with linear predictor $\eta = G\gamma + Fb$ with log-link (i.e., $\log(\phi_k) = \eta_k$). For the element $\tau \in \text{stack}(\gamma, b)$, the score equations of $p_{\beta,v}(h)$ are given by

$$\sum_{k=1}^{n} \left(\frac{\partial \phi_k}{\partial \eta_k} \right) \left(\frac{1-p_k}{2} \right) \left(\frac{d_k^* - \phi_k}{\phi_k^2} \right) + \frac{\partial l_3(\alpha; b)}{\partial \tau} = 0,$$

where $d_k^* = d_k \big/ (1-p_k)$ and $l_3(\alpha; b) = -\frac{m}{2}\log|2\pi\alpha| - \frac{1}{2\alpha} b^T b$. These are equivalnet to the score equations of a gamma generalized linear mixed model with response d^*, mean ϕ, log-link, linear predictor η and prior weight $(1-p)\big/2$. Similarly, $\hat{\lambda}$ can be obtained by fitting a gamma generalized linear model (Lee and Nelder, 2001; Lee *et al.*, 2006).

We can extend the conditional distribution of y_{ij} into the non-normal distribution. Consider a double hierarchical generalized linear model for y_{ij} satisfying

$$E(y_{ij}|v_i, b_i) = \mu_{ij} \text{ and } \text{var}(y_{ij}|v_i, b_i) = \phi_{ij} V(\mu_{ij}),$$

with the linear predictor

$$\eta_{ij} = g(\mu_{ij}) = X_{ij}^T \beta + Z_{ij}^T v_i, \tag{3.5}$$
$$v_i \sim N(0, \lambda_i),$$

and the dispersion model with log-link as shown in the model (3.3),

$$\log(\phi_{ij}) = G_{ij}^T\gamma + F_{ij}^T b_i,$$
$$b_i \sim N(0, \alpha I),$$

where $g(\cdot)$ is the link function, ϕ_{ij} is the dispersion parameter, and $V(\cdot)$ is the variance function specifying the part of the conditional variance of y_{ij} which depends on the conditional mean μ_{ij} ($i = 1, \cdots, m$; $j = 1, \cdots, q$).

The h-likelihood estimation for the double HGLM (3.5) can be viewed as that for an augmented generalized linear model (augmented GLM)) with the response $y_a =$ stack(y, ψ_M), where

$$E(y_a) = \mu_a = \begin{pmatrix} \mu \\ v \end{pmatrix} \text{ and var}(y_a) = \begin{pmatrix} \text{diag}\{\phi_{ij}V(\mu_{ij})\} & 0 \\ 0 & \text{diag}(\lambda_i) \end{pmatrix},$$

and an augmented linear predictor

$$\eta_a = g_a(\mu_a) = T\theta, \tag{3.6}$$

where $g_a(\cdot)$ is the link function such that $g_a(\mu_a) =$ stack$(g(\mu), v)$ and the augmented model matrix is

$$T = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix},$$

for the parameter $\theta =$ stack(β, v) (Lee and Nelder, 2001).

Estimating equations for $\hat{\beta}$ and $\hat{v}$ of the model (3.5) are those from the augmented GLM (3.6)

$$T^T\Sigma_a^{-1}T = T^T\Sigma_a^{-1}s_a,$$

where $s_a =$ stack$(s, 0)$ and $\Sigma_a = \begin{pmatrix} \Phi & 0 \\ 0 & \Lambda \end{pmatrix} W_a^{-1} = \begin{pmatrix} \Phi W^{-1} & 0 \\ 0 & \Lambda \end{pmatrix}$ with $\Phi =$ diag(ϕ_{ij}) and $\Lambda =$ diag(λ_i). The adjusted dependent variable s is a linearization of $g(y)$ around μ such that

$$s_{ij} = \eta_{ij} + (y_{ij} - \mu_{ij})\frac{\partial \eta_{ij}}{\partial \mu_{ij}},$$

and the weight matrix $W_a = \begin{pmatrix} W & 0 \\ 0 & I \end{pmatrix}$ contatins diagonal elements

$$w_{ij} = \left(\frac{\partial \mu_{ij}}{\partial \eta_{ij}}\right)^2 V(\mu_{ij})^{-1}.$$

In the same way with the normal model (3.3), let ϕ_k be the k-th diagonal elements of Φ and q_k be the k-th GLM leverage of the augmented model (3.6), and d_k be the k-th GLM deviance component such that

$$d = 2\int_{\hat{\mu}}^{y} \frac{(y-t)}{V(t)} dt,$$

then the score equations for the element $\tau \in \text{stack}(\gamma, b)$ are equivalent to the estimating equations of a gamma generalized linear mixed model with response $d^* = d/(1-q)$, mean ϕ, gamma error, log-link, linear predictor $\eta = \log(\phi) = G\gamma + Fb$, and prior weight $(1-q)/2$. Similarly, $\hat{\lambda}$ also can be obtained by fitting a gamma generalized linear model.

The corresponding effective degree of freedom of the model (3.5) is estimated by

$$p_D = tr(H^{-1}H^*)$$

where H and H^* are Hessian matrices for the h-likelihood and the conditional log-likelihood $l_1(\beta, \gamma; y|v, b)$ with repect to $(\beta, v, \gamma, b, \lambda)$. Because the conditional log-likelihood l_1 does not depend on the dispersion paramter λ, the Hessian matrix H^* is partitioned as

$$H^* = \begin{bmatrix} H^*_{\delta\delta} & 0 \\ 0 & 0 \end{bmatrix},$$

where $H^*_{\delta\delta} = -\partial^2 l_1 / \partial\delta\partial\delta^T$ for $\delta = \text{stack}(\beta, v, \gamma, b)$. In the same way, the Hessian matix H is also partitioned as

$$H = \begin{bmatrix} H_{\delta\delta} & H_{\delta\lambda} \\ H_{\lambda\delta} & H_{\lambda\lambda} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 h}{\partial\delta\partial\delta^T} & -\frac{\partial^2 h}{\partial\delta\partial\lambda^T} \\ -\frac{\partial^2 h}{\partial\lambda\partial\delta^T} & -\frac{\partial^2 h}{\partial\lambda\partial\lambda^T} \end{bmatrix}. \tag{3.7}$$

The inversion formula implies

$$p_D = tr(H^{-1}H^*) = tr\left\{(H_{\delta\delta} - H_{\delta\lambda}H_{\lambda\lambda}^{-1}H_{\lambda\delta})^{-1}H_{\delta\delta}^*\right\}, \tag{3.8}$$

where $(H_{\delta\delta} - H_{\delta\lambda}H_{\lambda\lambda}^{-1}H_{\lambda\delta})^{-1}$ is the top-left submatrix of H^{-1}. If we assume that the orthogonality between mean and dispersion (i.e., $n^{-1}E(\partial^2 h/\partial\beta\partial\phi^T) \to 0$), then $n^{-1}E(H_{\lambda\delta}) \to 0$ and $p_D = tr(H^{-1}H^*) \to tr(H_{\delta\delta}^{-1}H_{\delta\delta}^*)$.

3.2.2 Conditional Akaike information criterion for double hierarchical generalized linear model

We now define the conditional Akaike information criterion (cAIC) for double hierarchical generalized linear model to be used in model selection. Assume the true conditional distribution of y is $f_1(y|u,a)$ and that u and a are the true random effects vector for mean and dispersion with distributions $f_2(u)$ and $f_3(a)$, respectively. The prediction dataset is y^* such that y^* and y are independent conditional on u and a, and from the same distribution $f_1(\cdot|u,a)$. In other words, y and y^* share the same random effects u and a, but differ in their error term.

Definition. *In double hierarchical generalized linear models which contain random effects both for mean and dispersion models, conditional Akaike information (cAI) is defined to be*

$$\begin{aligned} cAI &= -2E_{f(y,u,a)}E_{f(y^*|u,a)}\log g\{y^*|\hat{\beta}(y),\hat{v}(y),\hat{\gamma}(y),\hat{b}(y)\} \\ &= \int -2\log g\{y^*|\hat{\beta}(y),\hat{v}(y),\hat{\gamma}(y),\hat{b}(y)\}f_1(y^*|u,a)f(y,u,a)dy^*dyduda \end{aligned} \tag{3.9}$$

where $f(y,u,a) = f_1(y|u,a)f_2(u)f_3(a)$ is the true joint distribution of y, u and a. $\hat{\beta}(y)$ and $\hat{v}(y)$ are estimators of fixed effects β and random effects v for the mean model, and $\hat{\gamma}(y)$ and $\hat{b}(y)$ are those of fixed effects γ and random effects b for the dispersion model based on data y.

Theorem. *Assume the the data y are generated from the double hierarchical generalized linear model* (3.5). *Let $\hat{\beta}(y)$, $\hat{v}(y)$, $\hat{\gamma}(y)$, and $\hat{b}(y)$ are h-likelihood estimators for β,v, γ, and b, respectively. Under the conditions A1-A15 given in the Appendix and with the h-likelihood estimator $\hat{\alpha}$, an asymptotically unbiased esimator of the cAI in* (3.9) *is given by*

$$cAIC = -2\log g\left(y|\hat{\beta}(y),\hat{v}(y),\hat{\gamma}(y),\hat{b}(y)\right) + 2\hat{p}_h,$$

where $\hat{p}_h = p_D = tr\left\{(H_{\delta\delta} - H_{\delta\lambda}H_{\lambda\lambda}^{-1}H_{\lambda\delta})^{-1}H_{\delta\delta}^{*}\right\}$ is the corresponding effective degree of freedom in (3.8).

In the double HGLM (3.5), let $\hat{\zeta}$ be the h-likelihood estimator for $\zeta = \text{stack}(\beta, v, \gamma, b, \lambda)$. Then, the asymptotic variance-covariance matrix of $\hat{\zeta} - \zeta$ was shown by Lee and Kim (2019): under appropriate conditions, as $\min(m, q) \to \infty$

$$\text{var}[\sqrt{n}(\hat{\zeta} - \zeta)] \to \left\{\frac{I(\zeta)}{n}\right\}^{-1}, \tag{3.10}$$

where $I(\zeta)$ is the expected h-likelihood information matrix

$$I(\zeta) = E\left[-\frac{\partial^2 h}{\partial\zeta\partial\zeta^T}\right],$$

and is estimated by the observed h-likelihood information matrix

$$I(\hat{\zeta}) = -\frac{\partial^2 h}{\partial\zeta\partial\zeta^T}\bigg|_{\zeta=\hat{\zeta}} = H\big|_{\zeta=\hat{\zeta}}.$$

The variance-covariance matrix of $\hat{\zeta} - \zeta$ is also can be estimated by using $p_v(h)$ or $p_{\beta,v}(h)$ for dispersion components. For example, $-\partial^2 h/\partial\tau\partial\tau^T$ is replaced by $-\partial^2 p_v(h)/\partial\tau\partial\tau^T$ such that

$$H_{(l)} = \begin{bmatrix} H_{\theta\theta} & H_{\theta\tau} \\ H_{\tau\theta} & H_{(l)\tau\tau} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 h}{\partial\theta\partial\theta^T} & -\frac{\partial^2 h}{\partial\theta\partial\tau^T} \\ -\frac{\partial^2 h}{\partial\tau\partial\theta^T} & -\frac{\partial^2 p_v(h)}{\partial\tau\partial\tau^T} \end{bmatrix},$$

where $\theta = \text{stack}(\beta, v)$ and $\tau = \text{stack}(\gamma, b, \lambda)$. If we use $H_{(l)}$ to estimate the bias correction term p, then we can obtain another bias correction estimator

$$\hat{p}_l = tr(H_{(l)}^{-1}H_1^{*}) = tr\left\{(H_{(l)\delta\delta} - H_{(l)\delta\lambda}H_{(l)\lambda\lambda}^{-1}H_{(l)\lambda\delta})^{-1}H_{\delta\delta}^{*}\right\}. \tag{3.11}$$

where $H_{(l)\delta\delta}$, $H_{(l)\delta\lambda}$, and $H_{(l)\lambda\lambda}$ are corresponding submatrix of $H_{(l)}$. Yu and Yau (2013) use this $H_{(l)}$ to compute the bias correction estimator $\hat{p}_{ml}$ in (3.2) with known ϕ (i.e., known γ and $b = 0$). $\hat{p}_l$ is the extension of $\hat{p}_{ml}$ considering the uncertainty in estimation of the dispersion ϕ_{ij}.

Let l be the marginal log-likelihood which is approximated by the adjusted profile likelihood $p_v(h)$

$$l = \log \int \exp(h) dv \approx p_v(h),$$

then the gradiant of l is

$$\frac{\partial l}{\partial \beta} = \frac{1}{\int \exp(h) dv} \int \frac{\partial h}{\partial \beta} \exp(h) dv.$$

If both the numerator and the denominator in the above equation are approximated by the first-order Laplace approximations, then

$$\frac{\partial l}{\partial \beta} \approx \frac{\frac{\partial h}{\partial \beta} \left| -\frac{1}{2\pi} \frac{\partial^2 h}{\partial v \partial v^T} \right|^{-1/2} \exp(h)}{\left| -\frac{1}{2\pi} \frac{\partial^2 h}{\partial v \partial v^T} \right|^{-1/2} \exp(h)} = \frac{\partial h}{\partial \beta}.$$

where v is evaluated at the solution of $\partial h / \partial v = 0$. This approximations imply that $H_{\theta\tau}$ in the partition of $H_{(l)}$ is approximately estimated by $-\partial^2 h / \partial\theta \partial\tau^T$ unlike Yu and Yau (2012) by $-\partial^2 h_a / \partial\theta \partial\tau^T$.

In the same way, if the right-bottom submatrix of H is replaced by $-\partial^2 p_{\beta,v}(h) / \partial\tau \partial\tau^T$ such that

$$H_{(r)} = \begin{bmatrix} H_{\theta\theta} & H_{\theta\tau} \\ H_{\tau\theta} & H_{(r)\tau\tau} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 h}{\partial\theta\partial\theta^T} & -\frac{\partial^2 h}{\partial\theta\partial\tau^T} \\ -\frac{\partial^2 h}{\partial\tau\partial\theta^T} & -\frac{\partial^2 p_{\beta,v}(h)}{\partial\tau\partial\tau^T} \end{bmatrix},$$

then the bias correction estimator $\hat{p}_r$ is obtained such that

$$\hat{p}_r = tr(H_{(r)}^{-1} H_{\delta\delta}^*) = tr\left\{ (H_{(r)\delta\delta} - H_{(r)\delta\lambda} H_{(r)\lambda\lambda}^{-1} H_{(r)\lambda\delta})^{-1} H_{\delta\delta}^* \right\}. \tag{3.12}$$

where $H_{(r)\delta\delta}$, $H_{(r)\delta\lambda}$, and $H_{(r)\lambda\lambda}$ are corresponding submatrix of $H_{(r)}$.

3.3 Numerical studies and applications

3.3.1 Simulations

Assessing the bias of cAIC, the estimator of cAI, is equivalent to examining the accuracy of bias correction estimators

- $\hat{p}_c = tr(H_1^{-1}H_1^*)$ in (3.1),
- $\hat{p}_h = tr(H^{-1}H^*) = tr\left\{(H_{\delta\delta} - H_{\delta\lambda}H_{\lambda\lambda}^{-1}H_{\lambda\delta})^{-1}H_{\delta\delta}^*\right\}$ in (3.8),
- $\hat{p}_l = tr(H_{(l)}^{-1}H^*) = tr\left\{(H_{(l)\delta\delta} - H_{(l)\delta\lambda}H_{(l)\lambda\lambda}^{-1}H_{(l)\lambda\delta})^{-1}H_{\delta\delta}^*\right\}$ in (3.11),
- $\hat{p}_r = tr(H_{(r)}^{-1}H^*) = tr\left\{(H_{(r)\delta\delta} - H_{(r)\delta\lambda}H_{(r)\lambda\lambda}^{-1}H_{(r)\lambda\delta})^{-1}H_{\delta\delta}^*\right\}$ in (3.12).

Simulations are conducted based on 1000 replications to evaluate their accuracies. We first consider two models:

$$\begin{aligned}
\textbf{Model 1} \quad : \quad & y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}, \\
& \qquad \text{with } v_i \sim N(0,\lambda) \text{ and } e_{ij} \sim N(0,\phi_{ij}), \\
& \text{and } \log(\phi_{ij}) = \gamma_0 + b_i, \quad \text{with } b_i \sim N(0,\alpha), \\
\textbf{Model 2} \quad : \quad & y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}, \\
& \qquad \text{with } v_i \sim N(0,\lambda) \text{ and } e_{ij} \sim N(0,\phi).
\end{aligned}$$

In order to generate the data from model 1 and 2, we set $\beta_0 = 1$, $\beta_1 = -0.5$, $x_{ij} \sim BI(1, 1/2)$, and $\lambda = 0.1$. We also set $\gamma_0 = -1$, $\alpha = 0.3$ for model 1, and $\phi = 0.135$ for model 2. In each setting, 1000 sets of y and y_* are generated. The true bias correction term p of cAIC are taken as the corresponding empirical means of $\log g\{y|\hat{\beta}(y), \hat{v}(y), \hat{\gamma}(y), \hat{b}(y)\} - \log g\{y^*|\hat{\beta}(y), \hat{v}(y), \hat{\gamma}(y), \hat{b}(y)\}$. We assume that the variance components are unknown, then plug-in method is used for $\hat{p}_c$.

In Table 3.1, decreasing relative biases are observed as the number of clusters and the cluster size increase, concurring that $\hat{p}_h$ is the asymptotically unbiased estimator of the bias correction term p. It is found that the relative bias of $\hat{p}_h$ is small, and as

a comparison, the relative bias of $\hat{p}_l$ and $\hat{p}_r$ also appear to be small. In general, $\hat{p}_h$ provides a more accurate estimation for p then $\hat{p}_c$ in different model settings, and $\hat{p}_h$, $\hat{p}_l$, and $\hat{p}_r$ are asmyptotically equivalent as q increases.

In addition to estimation of the accuracy, the model selection performance of the cAIC with $\hat{p}_h$ and other bias correction estimators are also investigated. We set $m = 10$ and $q = 20$. The model selection results among model 1 and model 2 are tabulated in Table 3.2. The results show that cAIC with $\hat{p}_h$ serves as a useful model selection tool for double HGLMs. The correct model identification percentage ranges from 90.8% to 93.0% for the proposed cAIC using $\hat{p}_h$. When the data are generated by the reduced model (model 2), consistent with Ha *et al.* (2007) and Greven and Kneib (2010), conventional cAIC with $\hat{p}_c$ tends to choose a more complex model. On the other hand, the correct model identification percentage of the proposed cAIC with $\hat{p}_h$ is much higher, indicating that the bias correction estimator $\hat{p}_h$ can accomodate mostly the degree of freedom loss in this situation.

Further simulation studies were conduncted to investigate the situations for different fixed effects and random effects in the mean model. We consider two additional models as below:

$$
\begin{aligned}
\textbf{Model 3} \quad : \quad & y_{ij} = \beta_0 + v_i + e_{ij}, \\
& \qquad \text{with } v_i \sim N(0, \lambda) \text{ and } e_{ij} \sim N(0, \phi_{ij}), \\
& \text{and } \log(\phi_i) = \gamma_0 + b_i, \quad \text{with } b_i \sim N(0, \alpha), \\
\textbf{Model 4} \quad : \quad & y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij}, \\
& \qquad \text{with } e_{ij} \sim N(0, \phi).
\end{aligned}
$$

Model 1 is the full model; Model 3 is the simplication of model 1 by assumming reduced fixed effects (i.e., $\beta_1 = 0$), and Model 4 is nested to model 2 by assumming null random effects (i.e., $\lambda = 0$). Simulation settings for model 3 and model 4 are same with the model 1 and model 2.

Table 3.3 and Table 3.4 show the model selection performance for different fixed

Table 3.1: Simulation results for the bias correction term p, and its estimators (1000 replications). Relative bias of $\hat{p}_c$, $\hat{p}_h$, $\hat{p}_l$, and $\hat{p}_r$ are displayed in parenthesses, and (m, q) represent number of clusters and cluster size.

Model	(m, q)	p	$\hat{p}_c$		$\hat{p}_h$		$\hat{p}_l$		$\hat{p}_r$	
1	(10, 10)	23.36	8.32	(-0.65)	17.67	(-0.24)	19.69	(-0.16)	19.99	(-0.14)
1	(10, 20)	22.54	9.28	(-0.59)	19.64	(-0.13)	19.51	(-0.13)	19.37	(-0.14)
1	(10, 40)	20.95	10.02	(-0.52)	20.32	(-0.03)	20.6	(-0.04)	20.02	(-0.04)
1	(40, 5)	34.88	24.45	(-0.30)	27.63	(-0.21)	38.00	(0.09)	37.02	(0.06)
1	(40, 10)	51.63	36.46	(-0.29)	49.92	(-0.03)	49.48	(-0.04)	49.43	(-0.04)
2	(10, 10)	10.67	9.64	(-0.11)	11.18	(0.05)	11.18	(0.05)	11.05	(0.04)
2	(10, 20)	11.38	10.27	(-0.10)	11.59	(0.02)	11.47	(0.01)	11.44	(0.01)
2	(10, 40)	11.55	10.62	(-0.08)	11.74	(0.01)	11.70	(0.01)	11.68	(0.01)
2	(40, 5)	35.14	32.31	(-0.08)	34.93	(-0.01)	33.99	(-0.03)	33.91	(-0.03)
2	(40, 10)	37.91	36.04	(-0.05)	37.45	(-0.01)	37.27	(-0.02)	37.26	(-0.02)

Table 3.2: Simulation results for the correct model identification percentages of the cAICs with the estimated bias correction factors (1000 replications).

	True model (%)	
Criteria	Model 1	Model 2
cAIC with $\hat{p}_c$	100	0.0
cAIC with $\hat{p}_h$	90.8	93.0
cAIC with $\hat{p}_l$	90.8	92.6
cAIC with $\hat{p}_r$	91.0	93.1

effects and random effects in the mean model. For the fixed effect model selection, the correct model identification percentage of all cAICs and mAIC are sufficiently high (Table 3.3). In the model 4, the fixed effects β and the dispersion parameters ϕ are orthogonal (i.e., $E(\partial^2 h/\partial\beta\partial\phi) = 0$), then conventional cAIC with $\hat{p}_c$ selects the correct model to the 87.8%, whereas correct model selecion rates of other cAICs and rAIC are higher.

The results of Table 3.5 suggest that the cAIC with $\hat{p}_h$ is a resonable criterion for double HGLM selection. cAIC with $\hat{p}_h$ seems to have a weak tendency towards more complex models and the bias correction estimators $\hat{p}_h$, $\hat{p}_l$, and $\hat{p}_r$ are comparable. In the following subsection, cAIC with $\hat{p}_h$ is employed as the model selection tool in real data applications : tramadol data and fimasartan data in chapter 2.

3.3.2 Application: tramadol data

Two models are considered for the tramadol data in chapter 2: the bivariate hierarchical generalized linear model (bivariate HGLM) and independent linear mixed models. The parameter estimates and cAIC are displayed in Table 3.6. The biavariate HGLM has smaller cAIC=-75.14 than cAIC=-66.64 of independent models. This is same with the result using likelihood ratio test based on the restricted likelihood $p_{\beta,v}(h)$: selecting bivariate HGLM.

3.3.3 Application: Fimasartan data

Recall four models for fimasartan data in chapter 2 as follows:

M1 : bivariate double HGLM with a dispersion model

$$\log(\phi_{1ij}) = \gamma_{10} + b_{1i}, \;\; b_{1i} \sim N(0, \alpha_1).$$

Table 3.3: Simulation results for the correct model identification percentages of the cAICs with the estimated bias correction factors and mAIC (1000 replications).

	True model (%)	
Criteria	Model 1	Model 3
cAIC with $\hat{p}_c$	100	92.8
cAIC with $\hat{p}_h$	100	92.3
cAIC with $\hat{p}_l$	100	91.8
cAIC with $\hat{p}_r$	100	92.0
mAIC	99.7	99.9

Table 3.4: Simulation results for the correct model identification percentages of the cAICs with the estimated bias correction factors and rAIC (1000 replications).

	True model (%)	
Criteria	Model 2	Model 4
cAIC with $\hat{p}_c$	100	87.8
cAIC with $\hat{p}_h$	99.6	98.8
cAIC with $\hat{p}_l$	99.8	95.1
cAIC with $\hat{p}_r$	99.7	97.7
rAIC	100	94.4

Table 3.5: Simulation results for the correct model identification percentages of the cAICs with the estimated bias correction factors (1000 replications).

	True model (%)			
Selected Model	Model 1	Model 2	Model 3	Model 4
cAIC with $\hat{p}_c$				
Model1	**100**	100	7.2	99.9
Model2	0.0	**0.0**	0.0	0.0
Model3	0.0	0.0	**92.8**	0.0
Model4	0.0	0.0	0.0	**1.0**
cAIC with $\hat{p}_h$				
Model1	**90.1**	7.0	6.6	5.4
Model2	8.8	**92.6**	8.4	1.1
Model3	0.0	0.0	**83.6**	0.0
Model4	1.1	0.4	1.5	**93.6**
cAIC with $\hat{p}_l$				
Model1	**90.3**	7.4	6.9	6.3
Model2	8.8	**92.4**	8.1	4.1
Model3	0.0	0.0	**83.8**	0.0
Model4	0.9	0.2	1.3	**89.5**
cAIC with $\hat{p}_r$				
Model1	**90.4**	6.9	6.7	6.1
Model2	8.6	**92.8**	7.6	2.3
Model3	0.0	0.0	**84.5**	0.0
Model4	1.0	0.3	1.3	**91.8**

Table 3.6: Results of analysis for the tramadol data.

		bivariate HGLM				Independent models ($\rho = 0$)			
		Estimate	S.E.	z-score	P-value	Estimate	S.E.	z-score	P-value
C_{max}	β_{10}	5.965	0.176	33.841	<0.001	5.965	0.154	38.632	<0.001
	β_{11}	-0.144	0.040	-3.598	<0.001	-0.144	0.046	-3.139	0.002
	β_{12}	-0.009	0.040	-0.224	0.823	-0.009	0.046	-0.195	0.845
	β_{13}	-0.066	0.104	-0.635	0.526	-0.066	0.086	-0.766	0.444
	ϕ_1	0.010				0.013			
AUC	β_{20}	8.158	0.130	62.630	<0.001	8.158	0.158	51.711	<0.001
	β_{21}	-0.054	0.022	-2.470	0.014	-0.054	0.022	-2.470	0.014
	β_{22}	-0.040	0.022	-1.844	0.065	-0.040	0.022	-1.844	0.065
	β_{23}	-0.120	0.079	-1.506	0.132	-0.120	0.097	-1.229	0.219
	ϕ_2	0.003				0.003			
	λ_1	0.028				0.016			
	λ_2	0.017				0.017			
	ρ	0.884				0			
	cAIC	-75.14				-66.64			

M2 : independent double HGLM with a dispersion model

$$\log(\phi_{1ij}) = \gamma_{10} + b_{1i}, \;\; b_{1i} \sim N(0, \alpha_1),$$

where $\rho = 0$.

M3 : bivariate HGLM with $\phi_{1ij} = \phi_1$.

M4 : independent linear mixed models where $\rho = 0$ and $\phi_{1ij} = \phi_1$.

M1 is the full model and the others are various simplications of it by assuming null components, i.e. M2 ($\rho = 0$), M3 ($\alpha_1 = 0$), M4 ($\rho = 0$, $\alpha_1 = 0$). Estimating results and cAIC are reported in Table 3.7. cAIC selects M1 as the best fitting model with smallest cAIC=103.99.

Table 3.7: Results of analysis for the fimasartan data.

		M1: bivariate double HGLM				M2: independent double HGLMs			
		Estimate	S.E.	z-score	P-value	Estimate	S.E.	z-score	P-value
C_{max}	β_{10}	5.299	0.146	36.372	<0.001	5.301	0.160	33.193	<0.001
	β_{11}	0.088	0.173	0.506	0.613	0.086	0.203	0.425	0.671
	$\exp(\gamma_{10})$	0.377				0.413			
	α_1	0.142				0.120			
AUC	β_{20}	6.633	0.079	83.493	<0.001	6.633	0.090	73.682	<0.001
	β_{21}	0.151	0.082	1.836	0.066	0.151	0.086	1.751	0.080
	ϕ_2	0.075				0.071			
	λ_1	0.135				0.092			
	λ_2	0.075				0.083			
	ρ	0.831				0			
	cAIC	103.99				111.10			
		M3:bivariate HGLM				M4:independent LMM			
		Estimate	S.E.	z-score	P-value	Estimate	S.E.	z-score	P-value
C_{max}	β_{10}	5.291	0.155	34.157	<0.001	5.291	0.170	31.149	<0.001
	β_{11}	0.092	0.191	0.481	0.631	0.092	0.221	0.415	0.679
	ϕ_1	0.436				0.466			
AUC	β_{20}	6.633	0.079	83.493	<0.001	6.633	0.090	73.682	<0.001
	β_{21}	0.151	0.082	1.836	0.066	0.151	0.086	1.751	0.080
	ϕ_2	0.075				0.071			
	λ_1	0.122				0.082			
	λ_2	0.075				0.083			
	ρ	0.832							
	cAIC	106.51				113.32			

Chapter 4

Concluding remarks

In this thesis, we apply multivatiate double hierarchical generalized linear models to C_{max} and AUC, two co-primary endpoints for testing the bioequivalence. We can add other pharmacokinetic endpoints to the model such as T_{max}, clearance, and half-life, even pharmacodynamic endpoints. Furthermore, when the correlation between outcomes is of interest, it can be analyzed by this multivariate models.

We define conditional Akaike information for double hierarchical generalized linear models and propose its asymptotically unbiased estimator conditional Akaike information criterion. The estimated bias correction term is closely related to the effective degree of freedom of Ha *et al.* (2007). Simulations in current thesis are limited to normal models, but it can be extended to the non-normal models such as gamma double HGLMs.

Appendix: Conditions and proof of theorem

Let the data y be generated from the model (3.5), and $\hat{\beta}, \hat{v}, \hat{\gamma}, \hat{b}$, and $\hat{\lambda}$ be h-likelihood estimators. Note that (m, q) denotes the number of clusters and the cluster size and $n = mq$.

Let H and H^* be the Hessian of h-likelihood and conditional log-likelihood l_1 with respect to $(\beta, v, \gamma, b, \lambda)$, and $V = E_{y|v,b}(H)$ and $V^* = E_{y|v,b}(H^*)$ be their conditional expectations given random effects v, b. We assume that the following conditions hold:

A1 The true β, v, γ, b, and λ are unique, and are in the interior of a convex closed bounded set of parameter space.

A2 The fixed effects component $\hat{\beta}$ in the mean model satisfies $\hat{\beta} \to \beta$ almost surely as $m, q \to \infty$.

A3 The random effects component $\hat{v}$ in the mean model satisfies $\max_{i=1,\cdots,m} \|\hat{v}_i \to v_i\| \to 0$ almost surely as $m, q \to \infty$.

A4 The fixed effects component $\hat{\gamma}$ in the dispersion model satisfies $\hat{\gamma} \to \gamma$ almost surely as $m, q \to \infty$.

A5 The random effects component $\hat{b}$ in the dispersion model satisfies $\max_{i=1,\cdots,m} \|\hat{b}_i \to b_i\| \to 0$ almost surely as $m, q \to \infty$.

A6 The fixed effects component $\hat{\lambda}$ satisfies $\hat{\lambda} \to \lambda$ almost surely as $m, q \to \infty$.

A7 For any m, q the first and second derivatives of h-likelihood and their conditonal expectations given v, b are existed and continuous on the parameter space.

A8 The ratio $q/m \to \infty$ as $m, q \to \infty$.

A9 As $m, q \to \infty$, $(H - V)/n$ and $(H^* - V^*)/n$ converges almost surely to 0 uniformly on the parameter space.

A10 As $m, q \to \infty$, $\sqrt{n}(\hat{\beta} - \beta) \to N(0, \Sigma_1)$ in distribution, and $n\|\hat{\beta} - \beta\|^2$ is uniformly integrable.

A11 As $q \to \infty$, $\sqrt{q}(\hat{v}_i - v_i) \to N(0, \Sigma_{2i})$ in distribution uniformly over i, and $n\|\hat{v}_i - v_i\|^2$ is uniformly integrable for all i.

A12 As $m, q \to \infty$, $\sqrt{n}(\hat{\gamma} - \gamma) \to N(0, \Sigma_3)$ in distribution, and $n\|\hat{\gamma} - \gamma\|^2$ is uniformly integrable.

A13 As $q \to \infty$, $\sqrt{q}(\hat{b}_i - b_i) \to N(0, \Sigma_{4i})$ in distribution uniformly over i, and $n\|\hat{v}_i - v_i\|^2$ is uniformly integrable for all i.

A14 As $m, q \to \infty$, $\sqrt{n}(\hat{\lambda} - \lambda) \to N(0, \Sigma_5)$ in distribution, and $n\|\hat{\lambda} - \lambda\|^2$ is uniformly integrable.

A15 The quantity $\frac{1}{n}H$ and $\frac{1}{n}H^*$ are bounded for all $(\beta, v, \gamma, b, \lambda)$, m and q.

Under the generalized linear mixed model the distributional comvergences in A10 and A11 are established in Lee and Nelder (1996) and Nie (2007). The estimating equations for variance components γ, b and λ are equivalent to a gamma generalized linear mixed model and a gamma generalized linear model (Lee and Nelder 2001; Lee *et. al.* 2006), then A12-A14 can be established in the same way. A15 can be directly varified under specific models such as the double hierarchical generalized linear model (3.5), since the derivatices can be explicitly calculated.

proof of theorem. Because data y are generated from model (3.5), thus

$$\log f_1(y|u, a) = \log g(y|v, b) = l_1(\beta, \gamma; y|v, b) = l_1(y|\delta),$$

where $\delta = \text{stack}(\beta, v, \gamma, b)$.

Let H^* be the Hessian matrix for $l_1(y^*|\delta)$ with repect to δ. Taking the second order Taylor espansion of $l_1(y^*|\hat{\delta})$ at true δ , then

$$\begin{aligned} l_1(y^*|\hat{\delta}) &= l_1(y^*|\delta) + \frac{\partial l_1(y^*|\delta)}{\partial \delta^T}\Big|_{\delta=\delta}(\hat{\delta}-\delta) \\ &\quad -\frac{1}{2}tr\left[\frac{1}{n}H^*_{\delta\delta}\big|_{\check{\delta}}\, n(\hat{\delta}-\delta)(\hat{\delta}-\delta)^T\right], \end{aligned}$$

where

$$H^*_{\delta\delta}\big|_{\check{\delta}} = -\frac{\partial^2 l_1\,(y^*|\delta)}{\partial\delta\partial\delta^T}\Big|_{\delta=\check{\delta}},$$

and $\check{\delta} = c\delta + (1-c)\hat{\delta}$ for some $0 < c < 1$.

Under the regularity conditions A9 and A15 in Appendix,

$$\begin{aligned} l_1(y^*|\hat{\delta}) &= l_1(y^*|\delta) + \frac{\partial l_1(y^*|\delta)}{\partial \delta^T}\Big|_{\delta=\delta}(\hat{\delta}-\delta) \\ &\quad -\frac{1}{2}tr\left[V_2\, n(\hat{\delta}-\delta)(\hat{\delta}-\delta)^T\right] + o_p(1), \end{aligned}$$

where $V_2 = n^{-1}E_{v,b}E_{y|v,b}(H^*_{\delta\delta})$.

Under the regularity conditions A2-A14 in Appendix, following (3.10) and the inversion formula for the partitioned H in (3.7),

$$\text{var}\left[\sqrt{n}(\hat{\delta}-\delta)\right] = V_1^{-1} + o(1)$$

where $V_1 = n)-1E_{v,b}E_{y|v,b}(H_{\delta\delta}-H_{\delta\lambda}H^{-1}_{\lambda\lambda}H_{\lambda\delta})$. The matrix $(H_{\delta\delta}-H_{\delta\lambda}H^{-1}_{\lambda\lambda}H_{\lambda\delta})^{-1}$ is the top-left submatrix of H^{-1}.

Take expectation in both sides with respect to $f_1(y^*|v,b)f(y,v,b)$, then

$$\begin{aligned} E_{y,v,b}E_{y^*|v,b}\left\{l_1(y^*|\hat{\delta})\right\} &= E_{y^*,v,b}\left\{l_1(y^*|\delta)\right\} \\ &\quad -\frac{1}{2}tr\left[V_2E_{y,v,b}\left\{n(\hat{\delta}-\delta)(\hat{\delta}-\delta)^T\right\}\right] + o(1) \\ &= E_{y^*,v,b}\left\{l_1(y^*|\delta)\right\} - \frac{1}{2}tr\left(V_1^{-1}V_2\right) + o(1). \qquad (1) \end{aligned}$$

Let $\hat{\delta}_*$ be the h-likelihood estimator based on y^*, one can establish

$$\begin{aligned} l_1(y^*|\delta) &= l_1(y^*|\hat{\delta}_*) + \frac{\partial l_1(y^*|\delta)}{\partial \delta)^T}\Big|_{\delta=\hat{\delta}_*}(\delta-\hat{\delta}_*) \\ &\quad -\frac{1}{2}tr\left[\frac{1}{n}H^*\Big|_{\check{\delta}_*} n(\delta-\hat{\delta}_*)(\delta-\hat{\delta}_*)^T\right], \end{aligned}$$

where

$$H^*\big|_{\check{\delta}_*} = -\frac{\partial^2 l_1\left(y^*|\delta\right)}{\partial\delta\partial\delta^T}\bigg|_{\delta=\check{\delta}_*},$$

and $\check{\delta}_* = d\delta + (1-d)\hat{\delta}_*$ for some $0 < d < 1$.

Using score equations for $\hat{\delta}_*$,

$$\frac{\partial l_1\left(y^*|\delta\right)}{\partial\delta}\bigg|_{\delta=\hat{\delta}_*} - \begin{pmatrix} 0 \\ \hat{\Lambda}^{-1}\hat{v}_* \\ -\frac{1}{2}tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial\gamma_k}\right\}_{(\gamma_k\in\gamma)} \\ -\frac{1}{2}tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial b_k}\right\}_{(b_k\in b)} + \frac{1}{\hat{\alpha}}\hat{b}_* \end{pmatrix} = 0,$$

where $P = T(T^T\Sigma_a^{-1}T)^{-1}T^T\Sigma_a^{-1}$ is the hat matrix of the augmented model (3.6)

$$tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial\gamma_k}\right\}_{(\gamma_k\in\gamma)} = \left(tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial\gamma_1}\right\}_{\delta=\hat{\delta}_*}, tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial\gamma_2}\right\}_{\delta=\hat{\delta}_*}, \cdots\right)^T$$

and

$$tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial b_k}\right\}_{(b_k\in b)} = \left(tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial b_1}\right\}_{\delta=\hat{\delta}_*}, tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial b_2}\right\}_{\delta=\hat{\delta}_*}, \cdots\right)^T,$$

it follows that

$$\begin{aligned} l_1(y^*|\delta) &= l_1(y^*|\hat{\delta}_*) + \hat{v}_*^T\Lambda^{-1}(v-\hat{v}_*) - \frac{1}{2}tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial\gamma_k}\right\}^T_{(\gamma_k\in\gamma)}(\gamma-\hat{\gamma}_*) \\ &\quad + \left[-\frac{1}{2}tr\left\{P\frac{\partial(\log\Sigma_a)}{\partial b_k}\right\}^T_{(b_k\in b)} + \frac{1}{\hat{\alpha}}\hat{b}_*\right]^T(b-\hat{b}_*) \\ &\quad -\frac{1}{2}tr\left[V_2\, n(\hat{\delta}_*-\delta)(\hat{\delta}_*-\delta)^T\right] + o_p(1). \end{aligned}$$

Take expectation in both sides with respect to $f(y^*, v, b)$, the regularity conditions A3-A5, A9-A15 in Appendix imply that

$$\begin{aligned} E_{y^*,v,b}\left\{l_1(y^*|\delta)\right\} &= E_{y^*,v,b}\left\{l_1(y^*|\hat{\delta}_*)\right\} \\ &\quad -\frac{1}{2}tr\left[V_2 E_{y^*,v,b}\left\{n(\hat{\delta}_*-\delta)(\hat{\delta}_*-\delta)^T\right\}\right] + o(1) \\ &= E_{y^*}\left\{l_1(y^*|\hat{\delta}_*)\right\} - \frac{1}{2}tr\left(V_1^{-1}V_2\right) + o(1). \end{aligned} \tag{2}$$

Substitute $E_{y^*,v,b}\left\{l_1(y^*|\delta)\right\}$ in (1) by the right-hand-side of (2), as y and y^* are identical in distribution, it follows that

$$E_{y,v,b}E_{y^*|v,b}\left\{l_1(y^*|\hat{\delta})\right\} = E_y\left\{l_1(y|\hat{\delta})\right\} - tr\left(V_1^{-1}V_2\right) + o(1).$$

Because $\|n^{-1}(H_{\delta\delta} - H_{\delta\lambda}H_{\lambda\lambda}^{-1}H_{\lambda\delta})|_{\delta=\hat{\delta}} - V_1\| = o_p(1)$ and $\|n^{-1}H^*_{\delta\delta}|_{\delta=\hat{\delta}} - V_2\| = o_p(1)$, then

$$tr\left(V_1^{-1}V_2\right) - tr\left\{(H_{\delta\delta} - H_{\delta\lambda}H_{\lambda\lambda}^{-1}H_{\lambda\delta})^{-1}H^*_{\delta\delta}\right\}|_{\delta=\hat{\delta}} = o_p(1).$$

The last equation concludes the proof. □

Bibliography

[1] Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *Selected Papers of Hirotugu Akaike* (pp. 199–213). New York, NY: Springer New York.

[2] Berger, R. L., and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, **11**(4), 283-319.

[3] Burnham, K. P., and Anderson, D. R. (2003). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*: Springer New York.

[4] Donohue, M. C., Overholser, R., Xu, R., and Vaida, F. (2011). Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika*, **98**(3), 685-700.

[5] Greven, S., and Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, **97**(4), 773-789.

[6] Ha, I. D., Lee, Y., and MacKenzie, G. (2007). Model selection for multi-component frailty models. *Statistics in Medicine*, **26**(26), 4790-4807.

[7] Hodges, J. S., and Sargent, D. J. (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika*, **88**(2), 367-379.

[8] Lee, Y., and Kim, G. (2019). Properties of h-likelihood estimators in clustered data. *to appear at International Statistical Review*.

[9] Lee, Y., and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B-Methodological*, **58**(4), 619-656.

[10] Lee, Y., and Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, **88**(4), 987-1006.

[11] Lee, Y., and Nelder, J. A. (2006). Double hierarchical generalized linear models. *Journal of the Royal Statistical Society Series C-Applied Statistics*, **55**, 139-167.

[12] Lee, Y., Nelder, J. A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*: CRC Press.

[13] Lee, Y., Nelder, J. A., and Pawitan, Y. (2017). *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood (2nd Edition)*: CRC Press.

[14] Lee, Y., and Noh, M. (2012). Modelling random effect variance with double hierarchical generalized linear models,*Statistical Modelling*, **12**(6), 487-502.

[15] Lee, Y., Ronnegard, L., and Noh, M. (2017). *Data analysis using hierarchical generalized linear models with R*: Chapman and Hall/CRC.

[16] Molas, M., Noh, M., Lee, Y., and Lesaffre, E. (2013). Joint hierarchical generalized linear models with multivariate Gaussian random effects. *Computational Statistics and Data Analysis*, **68**, 239-250.

[17] Nie, L. (2007). Convergence rate of MLE in generalized linear and nonlinear mixed-effects models: Theory and applications. *Journal of Statistical Planning and Inference*, **137**(6), 1787-1804.

[18] Spiegelhalter, D. J., Best, N. G., Carlin, B. R., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **64**, 583-616.

[19] Vaida, F., and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**(2), 351-370.

[20] Ye, J. (1998). On Measuring and Correcting the Effects of Data Mining and Model Selection. *Journal of the American Statistical Association*, **93**(441), 120-131.

[21] Yi, S., Chung, Y. J., Kim, T. E., Shin, H. S., Yoon, S. H., Cho, J. Y., Jang, I. J., Shin,S. G., and Yu, K. S. (2011). Pharmacokinetics of extended-release versus conventional tramadol/acetaminophen fixed-dose combination tablets: an open-label, 2-treatment, multiple-dose, randomized-sequence crossover study in healthy korean male volunteers. *Clinical Therapeutics*, **33**(6), 728-737.

[22] Yi, S., Kim, T. E., Yoon, S. H., Cho, J. Y., Shin, S. G., Jang, I. J., and Yu, K. S. (2011). Pharmacokinetic interaction of fimasartan, a new angiotensin II receptor antagonist, with amlodipine in healthy volunteers. *Journal of Cardiovascular Pharmacology*, **57**(6), 682-689.

[23] Yu, D., and Yau, K. K. W. (2012). Conditional Akaike information criterion for generalized linear mixed models, *Computational Statistics and Data Analysis*, **56**, 629-644.

초 록

1996년 Lee와 Nelder가 제안한 다단계우도 (H-likelihood) 방법은 다양한 데이터의 분석에 사용되고 있다. 특히 클러스터 안에서 반복 측정된 데이터는 다단계 일반화 선형모형(HGLM)을 통하여 분석할 수 있다. 상관관계가 있는 다중 자료를 분석하고자 할 때는 다변량 이중 다단계 일반화 선형모형 (multivariate double HGLM)을 고려할 수 있다.

이 논문은 시험제품과 대조제품사이의 약동학적 유사성을 평가하는 생물학적 동등성 검정에 다변량 이중 다단계 일반화 선형모형을 적용하였다. 만약 대조제품 대비 시험제품의 AUC와 C_{max}의 기하평균비의 90% 신뢰구간이 생물학적 동등성 마진인 $(0.8,\ 1.25)$ 범위에 포함된다면, 시험제품은 생물학적으로 동등하다고 판단된다. 두 가지 일차변수인 AUC와 C_{max} 사이에 서로 강한 상관관계가 있다면, 다변량 이중 다단계 일반화 선형모형을 이용할 때 추정된 처치 효과에 대한 표준오차가 더 작아지고, 따라서 기하평균비의 90% 신뢰구간이 더 좁아지는 결과를 가져온다.

다양한 모델 중에서 최적합 모델을 선택하기 위하여, 우리는 이중 다단계 일반화 선형모형에 대한 conditional Akaike information(cAI)을 정의하고, 이중 다단계 일반화 선형모형의 effective degree of freedom을 이용하여 cAI에 대한 점근적 불편 추정량인 contional Akaike informaiton criterion (cAIC)를 제안하였다. 기존의 cAIC와 이 논문에서 제안된 cAIC의 정확도와 최적 모델 선택 수행력을 비교하고, 이를 실제 데이터에 적용하여 최적모델 선택을 수행하였다.

주요어: 다단계 우도, 다단계 일반화 선형모형, 임상 약리 자료, 생물학적 동등성 시험, 자유도, 모델 선택, 아카이케 인포메이션
학번: 2007-20255