DETECTING INFLUENTIAL OBSERVATIONS AND SUBJECTS IN GENERALIZED LINEAR MIXED MODELS

Mario J.N.M. Ouwens, Frans E.S. Tan and Martijn P.F. Berger
Maastricht University, Department of Methodology and Statistics, the Netherlands

Abstract

This paper discusses the generalization of the subject-oriented local influence measures for normally distributed responses to observation-oriented influence measures for generalized linear models with random effects. A two-step diagnostic procedure is proposed. The first step is to search for influential subjects. A search for influential observations is proposed as second step. An illustration of a two-treatment multiple period cross-over trial demonstrates the practical importance of the detection of influential observations in addition to the detection of influential subjects.

Keywords: Influential Observations, Influential Subjects, Local Influence, GLMM, Poisson regression, Random effects, GLM, Likelihood Displacement.

Correspondence to: Mario J.N.M. Ouwens, Maastricht University, Department of Methodology and Statistics, PO Box 616, 6200 MD Maastricht, the Netherlands. Phone: +31-43-3882277, e-mail: mario.ouwens@stat.unimaas.nl

Local influence for generalized linear mixed effects models

1. Introduction

The interest in generalized linear mixed models has grown steadily during the past decades (Diggle, Liang and Zeger, 1994). Unfortunately, the estimates of the model parameters may heavily depend on a small part of the dataset or even on one particular observation or subject. For ease of presentation, a set of observations or subjects is called a data structure. A data structure which has a large impact on the estimated model parameters will be called influential. An approach to detect influential data structures is to compare the estimates of the model parameters based on the sample with and without the data structures of interest. Diagnostics based on this approach are called deletion diagnostics and are discussed for normally distributed responses by Christensen, Pearson and Johnson (1992), Bannerjee and Frees (1997) and Tan, Ouwens and Berger (1999). The only influence diagnostics for generalized linear mixed models discussed in literature are the deletion diagnostics of Preisser and Qaqish (1996) in the context of marginal models. For these models the average of the responses for each covariate pattern is modeled.

A second approach to detect influential data structures is based on the curvature of the log-likelihood function. One of the measures based on this approach is local influence and has been developed by Cook (1986). Local influence for generalized linear models is discussed by Thomas and Cook (1989, 1990). Although local influence for longitudinal models with normally distributed responses has already been discussed by Lesaffre and Verbeke (1998), their discussion is limited to the subject level. In this paper local influence will be generalized to the observational level.

Influential observations and subjects have disproportionally large local influence values. Consequently, to detect influential observations and influential subjects we have to compare the influence values with each other. Using this approach, the detection of influential observations is also important. In the dataset used in this paper, for example, the influential observations are distributed across several subjects in such a way that.the subjects themselves were not detected to be influential. Another reason why the detection of influential observations is important, is that subject-oriented influence measures cannot discriminate between influential subjects due to subject specific characteristics and influential subjects due to influential observations within those subjects. Consequently, if the influential data structures were to be deleted without taking into account the source of the influence, this may lead to an unnecessary loss of information.

The detection of influential subjects remains important. Suppose, for example, that the subject specific parameters of a subject are disproportionally large, but the observations are fitted well by the estimate of the subject specific profile. In that case, the estimated subject specific profile will not change significantly due to the deletion of an arbitrary observation of that subject. Consequently, the subject will not be detected by the evaluation of the influence of observations one at a time. As a result both the detection of influential observations as well as the detection of influential subjects need to be accounted for. Therefore we propose to use a two-step procedure. In the first step the subject-oriented local influence measure is used. In the second step, the observation-oriented local influence measure is used. This procedure can be applied iteratively.

In section two the underlying regression model is specified. The observation-oriented local influence approach is proposed in section three. It is shown that the subject-oriented local

influence approach is a special case of the observation-oriented influence approach. Finally, a real dataset is analyzed in section four and conclusions are drawn.

2. Model specification

Conditional on the subject specific parameters, the responses y_{ij} of subject i, $i = 1,...,N$, at time point j, $j = 1,...,n_i$, are independent and drawn from an exponential density function (Diggle, Liang and Zeger, 1994, pg. 134) :

$$f(y_{ij}|\theta_{ij}) = \exp\left[(y_{ij}\theta_{ij} - \psi(\theta_{ij}))/a_{ij}(\phi) + c_{ij}(y_{ij},\phi)\right], \quad (1)$$

where θ_{ij} is the canonical form of the location parameter and is a function of the conditional mean μ_{ij}, $a_{ij}(\phi)$ is a known function of the possibly unknown dispersion parameter, or vector of dispersion parameters, ϕ and c_{ij} is a known function of the dispersion parameter and the responses (McCullagh and Nelder, pg 28). ψ is a known function, such that the conditional mean of y_{ij} is equal to $\mu_{ij} = E(y_{ij}|\theta_{ij}) = \partial\psi(\theta_{ij})/\partial(\theta_{ij})$. The subjects are assumed to be independent and the subject specific parameters $\boldsymbol{b}_i$ are outcomes of the normal distribution $\Phi(0,G)$ where G is a $q \times q$ variance-covariance matrix. It should be noted, that the dispersion parameter ϕ may be unknown to incorporate the normal density function.

The Poisson density used in the real data example in this paper is a member of the family of density functions of the form (1). The Poisson density is equal to

$$f(y_{ij}|\mu_{ij}) = \frac{\mu_{ij}^{y_{ij}}\exp(-\mu_{ij})}{y_{ij}!} = \exp[y_{ij}\log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)] \quad (2)$$

and can be obtained from expression (1) by taking $a_{ij}(\phi) = 1$, $\theta_{ij} = \log(\mu_{ij})$, $\psi(\theta_{ij}) = \exp(\theta_{ij})$ and $c(y_{ij}, \phi) = -\log(y_{ij}!)$. In the real data analysis it is assumed that μ_{ij} is equal to $t_{ij}\exp(\boldsymbol{x}_{ij}'\beta + \boldsymbol{z}_{ij}'\boldsymbol{b}_i)$, where $\boldsymbol{x}_{ij}$ is the p design vector of the fixed effects of subject i at time point j and $\boldsymbol{z}_{ij}$ is the corresponding q design vector of the random effects. β is the p vector of fixed regression parameters and $\boldsymbol{b}_i$ is the vector of random effect regression parameters. t_{ij} is the number of days in the period, called the follow-up time. The model parameters can be estimated using the MIXPREG (source http://www.uic.edu/~hedeker/mixdos.html). In the next section we will discuss methods to detect influential observations and subjects.

3. Local influence

The first part of this section focusses on the detection of influential observations. The relationship between the observation-oriented and the subject-oriented local influence is the subject of the second part.

Some specific notation and definitions are used throughout this paper. The vector of fixed model parameters is denoted by ζ. For the Poisson model ζ consists of the regression parameters β and the q (q+1)/2 parameters specifying the variance-covariance matrix G of the random effects. The influence of a response is the influence of the observation determined by

the combination of that response and the corresponding design vectors of the random and fixed effects. In this paper it is assumed that the deletion of a certain response also leads to the deletion of the corresponding design vectors of the random and fixed effects. The complete sample is the sample consisting of all measured observations. The vector of responses $\boldsymbol{y}$ can be written as $[\boldsymbol{y}_1,\ldots\boldsymbol{y}_N]$, where $\boldsymbol{y}_i$ is the vector of n_i responses of subject i, $i = 1,\ldots, N$. Finally, let M_i be the set of responses to be evaluated. Then $\boldsymbol{y}_{iM_i}$ is the vector formed by stacking the responses in set M_i and $\boldsymbol{y}_{i(M_i)}$ is the vector formed by stacking the responses of subject i not in M_i.

3.1. Observation-Oriented Local Influence

Observation-oriented local influence measures for random effects models have not yet been discussed in the literature. In this section we will derive such a measure using the approach to evaluate the influence of a data structure by comparing the maximum likelihood estimates based on the sample with and without that data structure. The estimates can be compared using the Likelihood Displacement given by Cook and Weisberg (1982):
where $\hat{\zeta}$ and $\hat{\zeta}_{(M_i)}$ are the maximum likelihood estimates based on the complete sample with

$$LD_{M_i}=2[L(\boldsymbol{y}|\hat{\zeta})-L(\boldsymbol{y}|\hat{\zeta}_{(M_i)})], \tag{3}$$

and without the observations in M_i, respectively, and L denotes the log-likelihood function of the complete sample. Expression (3) can be Taylored using weighted log-likelihood functions. Cook (1986) called this the local influence approach. Since the subjects are independent, the log-likelihood function is the sum of the contributions of the individual subjects to the log-likelihood function. Let $l_i(\boldsymbol{y}_i|\zeta)$ and $l_{i(M_i)}(\boldsymbol{y}_{i(M_i)}|\zeta)$ be the contributions of the i-th subject to the log-likelihood functions of the complete sample and of the sample without the responses of subject i in M_i, respectively. The log-likelihood functions of the complete sample with and without the responses of subject i in M_i are both of the form

$$L(\boldsymbol{y}|\zeta,\omega_{M_i})=L(\boldsymbol{y}|\zeta)+(1-\omega_{M_i})(l_{i(M_i)}(\boldsymbol{y}_{i(M_i)}|\zeta)-l_i(\boldsymbol{y}_i|\zeta)), \tag{4}$$

where ω_{M_i} is a scalar. If $\omega_{M_i}=1$, expression (4) is equal to the log-likelihood function of the complete sample; if $\omega_{M_i}=0$, it is equal to the log-likelihood function of the complete sample without the responses in M_i. Expression (4) indicates that the weighted log-likelihood function can be defined as:

$$L(\boldsymbol{y}|\zeta,\boldsymbol{\omega})=L(\boldsymbol{y}|\zeta)+\sum_{i=1}^{N}\sum_{M_i}(1-\omega_{M_i})(l_{i(M_i)}(\boldsymbol{y}_{i(M_i)}|\zeta)-l_i(\boldsymbol{y}_i|\zeta)), \tag{5}$$

where ω is the vector of weights ω_{M_i}. Using expression (5), expression (3) can be written as:

$$LD(\tilde{\omega})=2[L(y|\hat{\zeta}_{\omega_0},\omega_0)-L(y|\hat{\zeta}_{\tilde{\omega}},\omega_0)], \quad (6)$$

where $\tilde{\omega}$ denotes the vector of weights for which each element is equal to 1, except the weight corresponding to set M_i which is equal to 0, and ω_0 is the vector of weights for which each element is equal to 1. The corresponding maximum likelihood estimates of the model parameters are $\hat{\zeta}_{\tilde{\omega}}$ and $\hat{\zeta}_{\omega_0}$. The assessment of influence due to the deletion of units can be interpreted as perturbing the empirical distribution function. Weights between 0 and 1 can be used to asses smaller perturbations of the empirical distribution function. Different vectors ω may be of interest. The evaluation of all possible vectors ω of interest, however, may be very laborous. One way to reduce the amount of work is to approximate the likelihood displacement by the second order Taylor expansion around ω_0. The likelihood displacement in ω_0 is zero. $L(y|\hat{\zeta}_{\omega_0},\omega_0)$ is always larger than or equal to $L(y|\hat{\zeta}_{\tilde{\omega}},\omega_0)$. Consequently, the value of the first derivative in ω_0 is zero. Hence the second order Taylor expansion depends only on the second derivative. Let $\Delta = \frac{\partial^2 L(y|\zeta,\omega)}{\partial\zeta\partial\omega}$ and $\boldsymbol{d}$ be the direction in which the second derivative is evaluated. For example, to evaluate the influence of the responses in M_i the only element of $\boldsymbol{d}$ unequal to zero is the element corresponding with ω_{M_i}. The local influence measure in ω_0 in the direction $\boldsymbol{d}$ is defined as twice the absolute value of the second derivative in the direction $\boldsymbol{d}$. For example, if we are interested in the influence of unit i is equal to the vector for which only the ith element is unequal to zero. The ith element is equal to 1. Cook (1986) shows that the local influence measure is equal to:

$$C_{\boldsymbol{d}}=2|\boldsymbol{d}'\Delta'\left(\frac{\partial^2 L(y|\zeta)}{\partial\zeta^2}\right)^{-1}\Delta\boldsymbol{d}|, \quad (7)$$

evaluated in $\zeta=\hat{\zeta}$. Local influence measures for a particular subset of the model parameters can be obtained by the procedure proposed by Cook (1986).

To calculate the value of the local influence C_d, the information matrix and Δ must be computed. The information matrix of the model parameters is standard output of most computer programs and can be interpreted without knowing the exact analytical expression. Thus we only need to concentrate on Δ to compute the local influence measure in (7).

Let S_i and $S_{i(M_i)}$ be the contributions of subject i to the score function based on the complete sample with an without set M_i, respectively. The column of Δ corresponding to weight ω_{M_i} is then equal to S_i - $S_{i(M_i)}$. The general expressions for (S_i - $S_{i(M_i)}$) are given in appendix A. For the Poisson model, S_i and $S_{i(M_i)}$ can be derived using the information given in the model section. The model parameters in the Poisson model are the regression parameters β and the parameters specifying the variance-covariance matrix G of the random effects. Consequently, the difference in the score functions consists of the subvector of the difference in the score function for the regression parameters and the subvector of the difference in the score function for the parameters specifying the random effects variance-covariance matrix G.

The subvector corresponding with the regression parameters is given by:

$$X_i'(\boldsymbol{y}_i-\mu_i)-X_{i(M_i)}'(\boldsymbol{y}_{i(M_i)}-\mu_{i(M_i)}), \tag{8}$$

where X_i is the design matrix with jth row $\boldsymbol{x}_{ij}$ and $X_{i(M_i)}$ is the design matrix of the responses of subject i not in M_i. μ_i is the vector of empirical bayes estimates of the responses of subject i, based on all n_i observations of subject i, and $\mu_{i(M_i)}$ is the vector of empirical bayes estimates of the responses of subject i not in M_i, based on the responses not in M_i.

Assuming that the random effects variance-covariance matrix is unstructured, the parameters specifying the variance-covariance matrix of the random effects are the elements of the matrix G. The derivative of the weighted log-likelihood function with respect to the jk-th element of the variance-covariance matrix of the random effects G, G_{jk}, and ω_{M_i} for $j \neq k$ in $\theta=\hat{\theta}$ is equal to the sum of the jk-th element and the kj-th element of the matrix

$$\frac{1}{2}\hat{G}^{-1}\left(E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}_i,\hat{\zeta})-E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}_{i(M_i)},\hat{\zeta})\right)\hat{G}^{-1}, \tag{9}$$

where $E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}_i,\hat{\zeta})$ and $E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}_{i(M_i)},\hat{\zeta})$ are the empirical bayes estimates of the second moment of the random effects of subject i, based on the complete sample with and without the responses in set M_i, respectively. The derivative with respect to G_{jj} is equal to the jj-th element of this matrix (9).

It should be noted that the larger the difference in contribution of subject i to the score functions, the more the corresponding maximum likelihood estimates differ. To take the shape of the log-likelihood function into account, expression (7) normalizes the difference in contribution using the variance-covariance matrix of the model parameters .

Note also that the form of expression (7) is very similar to the form of Cook's Distance (Cook, 1977), where the amount of difference in regression parameters is evaluated with respect to the metric defined by the variance-covariance matrix of the estimated regression parameters. Expression (8) is a function of the residuals and the location in the design space where the responses are measured, which is similar to the diagnostics in ordinary least squares. In the next section the relation between the subject-oriented local influence measure and the observation-oriented local influence measure will be discussed.

3.2. Subject-Oriented Local Influence

For the subject-oriented local influence measure each set M_i will consist of all observations of subject i, implying that the function $l_{i(M_i)}(\boldsymbol{y}_{i(M_i)}|\zeta)$ will be equal to zero. Let S_i be the set of all observations of subject i, $i=1,\ldots,N$. If we are only interested in the influence of subjects, expression (5) can be rewritten as:

$$L(\boldsymbol{y}|\zeta)+\sum_{i=1}^{N}\sum_{M_i}(1-\omega_{M_i})(l_{i(M_i)}(\boldsymbol{y}_{i(M_i)}|\zeta)-l_i(\boldsymbol{y}_i|\zeta))= \tag{10}$$

Local influence for generalized linear mixed effects models

$$=L(\boldsymbol{y}|\zeta)+\sum_{i=1}^{N}(1-\omega_{S_i})(-l_i(\boldsymbol{y}_i|\zeta))=\sum_{i=1}^{N}\omega_{S_i}l_i(\boldsymbol{y}_i|\zeta),$$

which is equal to the expression given by Lesaffre and Verbeke (1998) for the subject-oriented local influence approach. Hence the subject-oriented local influence measure is a special case of the observation-oriented local influence measure (7). In the following section these influence measures will be applied in the analysis of a real dataset.

4. Data description

The dataset which will be used to illustrate the influence measures is reported by McKnight and Van Den Eeden (1993) and is obtained from a two treatment multiple period crossover trial in which the number of headaches per week is repeatedly measured for 27 patients. In the first period, each patient received the placebo. In the other four periods the patients received either the placebo (P) or the aspartame (A), in random order, using the double-blind crossover treatment design. To wash out the effects of the treatment of the foregoing periods the periods are separated by one day. Although most of the counts are based on periods of seven days, the number of treatment days in the periods varied. Hedeker (1999) showed that the sequence in which the placebo and aspartame is given did not matter significantly. Furthermore, he assumed only random intercepts. This formed our motivation to assume that the patients only differed in their intercepts and that the sequence in which the observations are taken was not important. Table 1 shows the dataset. The data are grouped according to the use of aspartame and the use of placebo. The last column shows the actual sequence in which the placebo and the aspartame were given. Almost all periods were seven day periods, but some periods were smaller. The number of days in the periods smaller than seven days are given within brackets. The asterisks in Table 1 indicate the observations and patients which will be detected to be influential.

We fitted the poisson regression model with random intercepts:

$$E(y_{ij}|\beta,b_i) = t_{ij}\exp(\beta_0 + \text{DrugAsp}_{ij}\,\beta_1 + b_{0i}), \qquad \textbf{(11)}$$

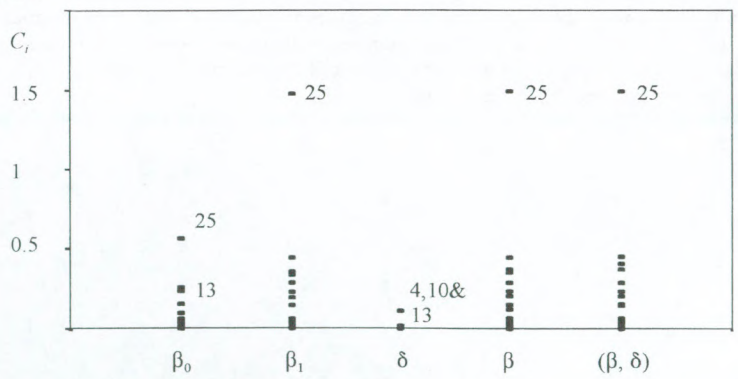
where β_0 is the fixed intercept, DrugAsp_{ij} indicates whether placebo (0) or aspartame (1) is given to patient i at timepoint j, β_1 is the corresponding fixed regression parameter, $\boldsymbol{b}_{0i}$ is the random intercept for patient i and t_{ij} is the number of days in the period. The standard deviation of the random intercept $\boldsymbol{b}_{0i}$ is denoted by δ.

Table 1. Number of Headaches of the Patients

Patient	Aspartame		Placebo			Sequence
1	0	0	3	3	1	PAPPA
2	5 *		0	2		PAP
3	2	2	2	3	2	PAPPA
4	0	0	0	0	0	PPAAP
5	0	0	3	2	0	PAPAP
6	1	1	1	0	3	PAPPA
7	4	3	1	1	2	PPAPA
8	0	1	1	1	1 (2)	PPAAP
9	2		2	0	1 (5)	PAPP
10	0	0	0	0	0	PPAPA
11	1	1 (3)	0	0	3	PPAPA
12	5 *	0	0	0	0	PPAAP
13	7	6	7	7	7	PAPAP
14	1	2	2	2	0	PPAPA
15	2	1	3	1	0	PPAAP
16	3		1	1		PAP
17	1 (1)		4			PA
18	1	0	0	1	1	PAPPA
19	1	0	0	1	1	PAPAP
20	1 (2)		1	6		PPA
21	2	6*	1	3	3	PAPPA
22	1	1	2	0	0	PAPPA
23	2	3	7 *	3	2	PAPAP
24	1	2	1	0	0	PPAPA
25*	6	7	1	1	0	PAPAP
26			0	1		PP
27	3	2 (4)	3	3	0	PPAPA

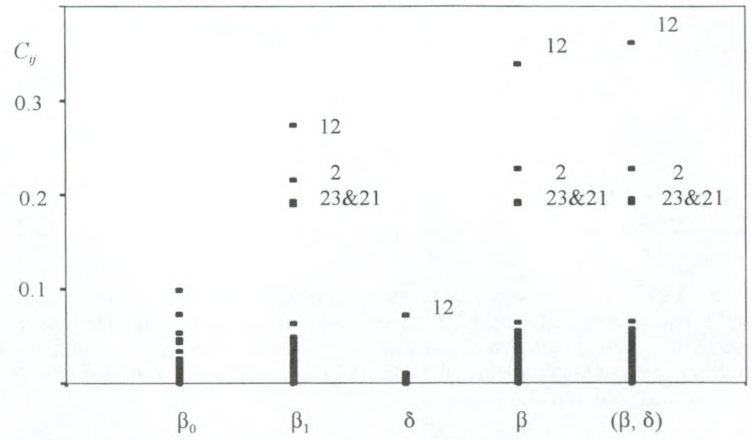
Figure 1 shows the subject-oriented local influence of the patients on the regression parameters β_0, β_1, the variance of the random intercept δ, the vector of regression parameters $\beta = (\beta_0, \beta_1)$ and the vector of all model parameters (β, δ), respectively. Patient 25 has a much larger influence on the regression parameters than the other patients of the sample. Table 1 shows that patient 25 has extremely many headaches using aspartame and a small number of headaches in the placebo condition.

Figure 1. Values of the subject oriented influence measures for five different sets of model parameters



After deletion of the observations of patient 25, the subject-oriented influence measure did not detect any other influential patients for the regression parameters. Figure 2 shows the influence at the observational level. The numbers in Figure 2 refer to the patients from which the influential observations are taken. The values of the influence measure for the four influential observations are more than three times larger than the values of the other observations. The influential observations are marked by an asterisk (*) in Table 1. It can be seen that the influential observations deviate from the subject-specific profiles. For example all observations of patient 12 are equal to 0, except the influential observation itself, which is equal to 5.

Figure 2. Values of the observation oriented local influence measures



To show the effect of the deletion of certain observations and/or patients on the estimated parameters, Table 2 displays the estimates of β_0, β_1 and δ, based on the sample omitting certain observations and/or patients.

Table 2. Estimated Model Parameters, Based on the Sample with and without Certain Patients or Observations.

	β_1	β_2	δ
sample	-1.72	0.28 (0.14)	0.69
Sample without subject 25	-1.70	0.15 (0.14)	0.70
Sample without subject 4,10,13 and 25	-1.61	0.19 (0.16)	0.39
Sample without subject 25 and without the influential observation of subject 12	-1.74	0.09 (0.15)	0.78
Sample without subject 25 and without the influential observation of subject 23	-1.72	0.20 (0.15)	0.68
Sample without subject 25 and without the 3 influential observations of the aspartame group	-1.74	-0.03 (0.16)	0.76
Sample without subject 25 and without the 4 influential observations	-1.77	0.02 (0.16)	0.74

From Table 2 it can be seen that the estimate of β_1 remains unstable after the deletion of patient 25 and that the estimate of β_1 varies between 0.20 and -0.03. The change in the estimate of β_1, due to the deletion of patient 25, is 0.13 and is as large as the additional change due to the deletion of the four influential observations. This means that the results still depend heavily on a small part of the dataset.

5. Discussion and Conclusions

In this paper the subject-oriented local influence measure proposed by Lesaffre and Verbeke (1998) is generalized to the observation-oriented local influence measure. Furthermore, local infuence is discussed for generalized linear models with random effects. The real data example shows that both the detection of influential observations and the detection of influential subjects are important. It shows that subjects may have a large influence while none of their observations is detected to be influential. On the other hand, the example shows also that influential observations may be distributed across subjects which are not detected to be influential by the subject-oriented influence measures. If we only use subject-oriented local influence measures the estimates of the parameters may heavily depend on the remaining small number of influential observations. Consequently, if one of these measures is not used, the possible instability of the parameter estimates may not be detected.

The influence is evaluated both for the whole set of parameters as well as for subsets of parameters. This can be motivated by the fact that some parameters may be of more interest than other parameters. Another motive is shown by the analyses of the real data example. The analyses of the real data example showed that if we had only used detection methods to detect

influential data structures for the whole set of parameters, some influential subjects for the estimation of the variance of the random effects were not detected to be influential. Unfortunately, the local influence approach is a local approach. The change in the estimated parameters due to the deletion of the detected subjects and observations may not always be disproportional large. This is especially the case for the detection of influential subjects for subsets of parameters. If, for example, the estimates of two regression parameters β_1 and β_2 are highly correlated, it may be expected that a change in the estimate of β_1 will be followed by a change in the estimate of β_2. This is taken into account by the multiplication with the variance-covariance matrix of the model parameters in expression (7). However, a change in β_1 may not necessarily be followed by a change in β_2. In other words, suppose that the influence of a subject for the estimate of β_1 is large, while the influence of that subject for β_2 is small. Then the local influence measure may detect that subject to be influential for the estimate of β_2, due to the large correlation between β_1 and β_2 and the large influence of that subject for the estimate of β_1. One method to avoid this is to orthogonalize the design matrices.

It should be emphasized that we used normally distributed random effects in this paper. Further research is needed to evaluate the robustness of the measures against violations of the normality assumption of the random effects. This paper discussed observation-oriented local influence and subject-oriented local influence in a very general form. Nevertheless, models assuming, for example, serial correlations are not contained in the discussed models. Fortunately, the weighted log-likelihood function in (5) can also be used for models with serial correlations. The relationship between local influence and goodness-of-fit seems to be very strong. Consequently, this is a topic for further research.

For the analyses of the real data example the random effects are assumed to be normally distributed. However, other distributions can be assumed. It should be noted that the only changes in the expressions if the distribution of the random effects is discrete rather than continuous is the change from integrands into sums.

6. References

Bannerjee, M. and Frees, E. W. (1997), Influence diagnostics for linear longitudinal models, *Journal of the American Statistical Association*, *92*, 999-1005.

Christensen, R., Pearson, L.M. and Johnson, W. (1992), Case deletion diagnostics for mixed models, *Techonometrics*, *34*, 38-45

Cook, R.D. (1977), Detection of influential observations in linear regression, *Technometrics*, *19*, 15-18.

Cook, R.D. (1986), Assessment of local influence, *Journal of the Royal Statistical Association*, *Series B*, *48*, 133-169.

Cook, R.D. , and Weisberg, S. (1982), *Residuals and influence in regression*, New York, Chapman and Hall.

Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994), *Analysis of longitudinal data*, Oxford, U.K.:Oxford Science Publications.

Hedeker, D. (1999), *Mixpreg manual Mixpcm.pdf*, http://www.uic.edu/~hedeker/mixdos.html

Lesaffre, E. and Verbeke, G. (1998), Local influence in linear mixed models, *Biometrics*, *54*, 570-582.

McCullagh, P. and Nelder, J.A. (1996), *Generalized linear models*, London, Chapman & Hall

McKnight, B., and van den Eeden, S.K. (1993), A Conditional Analyses for two-treatment multiple-period crossover designs with binomial or poisson outcomes and subjects who drop out, *Statistics in Medicine*, *12*, 825-834.

Preisser, J.S. and Qaqish, B.F. (1996), Deletion diagnostics for generalised estimating equations, *Biometrika*, *83*, 551-562.

Tan, F.E.S., Ouwens, M.J.N. and Berger, M.P.F. (1999), On the use of Cook's Distance for longitudinal regression models with random effects. Statistical modelling, Eds. H. Friedl, A. Berghold, G. Kamermann, proceeding of 14th International Workshop on Statistical Modelling, 686-690, Graz, Austria.

Thomas, W. and Cook, R.D. (1989), Assessing influence on regression coefficients in generalized linear models, *Biometrika*, *76*, 741-749.

Thomas, W. and Cook, R.D. (1990), Assessing influence on predictions from generalized linear models, *Technometrics*, *32*, 59-65.

Appendix A

The expression for the contributions of subject i to the score functions for the model parameters are given in this appendix.

If the conditional density of the responses of subject i, $i = 1,\ldots, N$ is given by $f_i \equiv \prod_{j=1}^{n_i} \exp\left(\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{a_{ij}(\phi)} + c_{ij}(y_{ij},\phi) \right)$ and the density of the random effects is given by $g = \frac{1}{\sqrt{(2\pi)^p |G|}} \exp(-\boldsymbol{b}_i' G^{-1} \boldsymbol{b}_i / 2)$, the contribution of subject i to the log-likelihood function based on the complete sample is equal to $\log \int f_i \; g \; db_i$. Note that the parameters specifying f_i are not used to specify g and vice versa, implying that the vector of model parameters ζ can be divided into the vector of parameters ζ_1, specifying G, and the vector of parameters ζ_2, specifying f_i. The derivative of the weighted log-likelihood function with respect to the jk-th element of the variance-covariance matrix of the random effects G, G_{jk}, and ω_{M_i} for $j \neq k$ in $\theta = \hat{\theta}$ is equal to the sum of the jk-th element and the kj-th element of the matrix

$$\frac{1}{2}\hat{G}^{-1}\left(E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}_i,\hat{\zeta}) - E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}_{i(M_i)},\hat{\zeta})\right)\hat{G}^{-1}, \qquad \textbf{(12)}$$

where $E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}_i,\hat{\zeta})$ and $E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}_{i(M_i)},\hat{\zeta})$ are the empirical bayes estimates of the second moment of the random effects of subject i, based on the complete sample with and without the responses in set M_i, respectively. The derivative with respect to G_{jj} is equal to the jj-th element of this matrix (9).

Let $S_i(\zeta_2)$ and $S_{i(M_i)}(\zeta_2)$ be the contributions of subject i to the score functions for ζ_2. $S_i(\zeta_2)$ is then equal to:

$$S_i(\zeta_2) = \int \sum_{j=1}^{n_i} \left(\frac{y_{ij} - \mu_{ij}}{a_{ij}(\phi)} \frac{\partial \theta_{ij}}{\partial \zeta_2} + \left(\frac{\partial c_{ij}(y_{ij},\phi)}{\partial \phi} - \frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{a_{ij}^2(\phi)} \frac{\partial a_{ij}(\phi)}{\partial \phi} \right) \frac{\partial \phi}{\partial \zeta_2} \right) \frac{f_i \; g}{\int f_i \; g \; db_i} db_i \qquad \textbf{(13)}$$

Consequently, $S_i(\zeta_2)$ is the empirical bayes estimate of

$$\sum_{j=1}^{n_i}\left(\frac{y_{ij}-\mu_{ij}}{a_{ij}(\phi)}\frac{\partial\theta_{ij}}{\partial\zeta_2}+\left(\frac{\partial c_{ij}(y_{ij},\phi)}{\partial\phi}-\frac{y_{ij}\theta_{ij}-\psi(\theta_{ij})}{a_{ij}^2(\phi)}\frac{\partial a_{ij}(\phi)}{\partial\phi}\right)\frac{\partial\phi}{\partial\zeta_2}\right). \qquad (14)$$

Analogously, $S_{i(M_i)}(\zeta_2)$ is the empirical bayes estimate of

$$\sum_{j:y_{ij}\notin M_i}\left(\frac{y_{ij}-\mu_{ij}}{a_{ij}(\phi)}\frac{\partial\theta_{ij}}{\partial\zeta_2}+\left(\frac{\partial c_{ij}(y_{ij},\phi)}{\partial\phi}-\frac{y_{ij}\theta_{ij}-\psi(\theta_{ij})}{a_{ij}^2(\phi)}\frac{\partial a_{ij}(\phi)}{\partial\phi}\right)\frac{\partial\phi}{\partial\zeta_2}\right), \qquad (15)$$

based on the responses not in M_i. It should be noted that if the dispersion ϕ is constant, the second term vanishes. This is the case when, for example, Poisson models and logistic models are used. In Appendix B. more attention will be paid to the Poisson model, because this regression model is used for the real data analyses.

Appendix B. Expression (8) to (10) for Poisson regression:

In this appendix the expressions (8) to (10) are formulated for Poisson regression with normally distributed random effects. To simplify the expressions we introduce some notations:

The location parameter: $w_{ik}=x_{ik}\beta+z_{ik}\boldsymbol{b}_i$

The conditional mean of the responses: $\lambda_{ik}=t_{ik}\exp(w_{ik})$

The Poisson density funtion: $f_{ij}=\exp(-\lambda_{ij}+y_{ij}\log t_{ij}+y_{ij}w_{ij}-\log(y_{ij}!))$

The Normal density function: $g_i=\dfrac{1}{(\sqrt{(2\pi)^p|G|})}\exp(-\boldsymbol{b}_i'G^{-1}\boldsymbol{b}_i/2)$

Expression (8): $X_i'(\boldsymbol{y}_i-\mu_i)-X_{i(M_i)}'(\boldsymbol{y}_{i(M_i)}-\mu_{i(M_i)})$

The matrices X_i and $X_{i(M_i)}$ and the vectors of responses $\boldsymbol{y}_i$ and $\boldsymbol{y}_{i(M_i)}$ are already defined in the text. The only vectors which depend on the choice of the model are μ_i and $\mu_{i(M_i)}$. For Poisson regression element k of μ_i is equal to

$$\int\lambda_{ik}\frac{\prod_{j=1}^{n_i}f_{ij}g_i}{\int\prod_{j=1}^{n_i}f_{ij}g_i d\boldsymbol{b}_i}d\boldsymbol{b}_i$$

evaluated in $\beta=\hat{\beta}$ and $G=\hat{G}$. Element k of $\mu_{i(M_i)}$ is for Poisson regression equal to

Local influence for generalized linear mixed effects models

$$\int \lambda_{lk} \frac{\prod_{j=1, j \notin M_i}^{n_i} f_{ij} g_i}{\int \prod_{j=1, j \notin M_i}^{n_i} f_{ij} g_i d\boldsymbol{b}_i} d\boldsymbol{b}_i$$

evaluated in $\beta = \hat{\beta}$ and $G = \hat{G}$.

Expression (9) : $\frac{1}{2}\hat{G}^{-1}\left(E(\boldsymbol{b}_i \boldsymbol{b}_i' | \boldsymbol{y}_i, \hat{\zeta}) - E(\boldsymbol{b}_i \boldsymbol{b}_i' | \boldsymbol{y}_{i(M_i)}, \hat{\zeta})\right)\hat{G}^{-1}$

The matrix $\hat{G}$ is already defined in the text. The conditional estimates of the second moment of the random effects are given by:

$$E(\boldsymbol{b}_i \boldsymbol{b}_i' | \boldsymbol{y}_i, \hat{\zeta}) = \int \boldsymbol{b}_i \boldsymbol{b}_i' \frac{\prod_{j=1}^{n_i} f_{ij} g_i}{\int \prod_{j=1}^{n_i} f_{ij} g_i d\boldsymbol{b}_i} d\boldsymbol{b}_i$$

and

$$E(\boldsymbol{b}_i \boldsymbol{b}_i' | \boldsymbol{y}_{i(M_i)}, \hat{\zeta}) = \int \boldsymbol{b}_i \boldsymbol{b}_i' \frac{\prod_{j=1, j \notin M_i}^{n_i} f_{ij} g_i}{\int \prod_{j=1, j \notin M_i}^{n_i} f_{ij} g_i d\boldsymbol{b}_i} d\boldsymbol{b}_i$$

both evaluated in $\beta = \hat{\beta}$ and $G = \hat{G}$.

Expression (10):

$$L(\boldsymbol{y}|\zeta) + \sum_{i=1}^{N} \sum_{M_i} (1 - \omega_{M_i})(l_{i(M_i)}(\boldsymbol{y}_{i(M_i)}|\zeta) - l_i(\boldsymbol{y}_i|\zeta)) =$$

The log-likelihood function $L(\boldsymbol{y}|\zeta)$ for Poisson regression is given by

$$\sum_{i=1}^{N} \log \int \prod_{j=1}^{n_i} f_{ij} g_i d\boldsymbol{b}_i$$

The contribution of subject i to the log-likelihood function based on all observations, $l_i(\boldsymbol{y}_i|\zeta)$, is equal to $\log \int \prod_{j=1}^{n_i} f_{ij} g_i d\boldsymbol{b}_i$ and the contribution of subject i to the log-likelihood function based

on all observations, except the observations in M_i, $l_{i(M_i)}(y_{i(M_i)}|\zeta)$, is equal to $\log\int\prod_{j=1,j\notin M_i}^{n_i} f_{ij}g_i d\boldsymbol{b}_i$.

Ontvangen: 20 augustus 1999
Geaccepteerd: 27 april 2000