Local Influence Diagnostics for Generalized Linear Mixed Models With Overdispersion

Journal:	*Journal of Applied Statistics*
Manuscript ID	CJAS-2015-0173.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	24-Feb-2016
Complete List of Authors:	Rakhmawati, Trias Wahyuni; I-BioStat, Universiteit Hasselt, Molenberghs, Geert; I-BioStat, Universiteit Hasselt, ; I-BioStat, Katholieke Universiteit Leuven, Verbeke, Geert; I-BioStat, Katholieke Universiteit Leuven, ; I-BioStat, Universiteit Hasselt, Faes, Christel; I-BioStat, Universiteit Hasselt,
Keywords:	Case deletion, Combined model, Logit-normal model, Poisson-normal model, Probit-normal model, Weibull-normal model
2010 Mathematics Subject Classification:	62F99, 62P10

To appear in the *Journal of Applied Statistics*
Vol. 00, No. 00, Month 20XX, 1–20

Local Influence Diagnostics for Generalized Linear Mixed Models With Overdispersion

Trias Wahyuni Rakhmawati[a∗] Geert Molenberghs[a,b] Geert Vebeke[b,a] Christel Faes[a,b]

[a] *I-BioStat, Universiteit Hasselt, B-3500 Hasselt, Belgium*
[b] *I-BioStat, KU Leuven, B-3000 Leuven, Belgium*

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

Since the seminal paper by Cook and Weisberg [9] , local influence, next to case deletion, has gained popularity as a tool to detect influential subjects and measurements for a variety of statistical models. For the linear mixed model the approach leads to easily interpretable and computationally convenient expressions, not only highlighting influential subjects, but also which aspect of their profile leads to undue influence on the model's fit [17]. Ouwens, Tan, and Berger [24] applied the method to the Poisson-normal generalized linear mixed model (GLMM). Given the model's non-linear structure, these authors did not derive interpretable components but rather focused on a graphical depiction of influence. In this paper, we consider GLMMs for binary, count, and time-to-event data, with the additional feature of accommodating overdispersion whenever necessary. For each situation, three approaches are considered, based on: (1) purely numerical derivations; (2) using a closed-form expression of the marginal likelihood function; and (3) using an integral representation of this likelihood. Unlike when case deletion is used, this leads to interpretable components, allowing not only to identify influential subjects, but also to study the cause thereof. The methodology is illustrated in case studies that range over the three data types mentioned.

Keywords: Case deletion; Combined model; Logit-normal model; Poisson-normal model; Probit-normal model; Weibull-normal model.

1. Introduction

Next to linear mixed models (LMM) for hierarchical Gaussian data [26], generalized linear mixed models (GLMM; [2, 19, 28]) have become a standard tool for the analysis of hierarchical data of a variety of data types. Routinely, after formulating and fitting a model, an assessment of model fit and a diagnostic analysis is advisable. Here, we are concerned with the detection of influential subjects.

A large variety of diagnostic tools is available for (generalized) linear models. Cook and Weisberg [9] and Chatterjee and Hadi [3] provide early treatises. In linear regression, Cook's distances [5–7] have been used extensively. They capture how much a parameter changes based on the contribution from one particular individual. If unduly large, the subject is considered influential. Linear mixed models, unlike linear models, generally do not allow for closed-form parameter estimators. Further, residual analysis is not straightforward, given the presence of both fixed- and random-effects, so that even uniquely defining residuals is not possible. For these and related reasons, Lesaffre and Verbeke [17] , chose local influence [1, 8] to examine influence in linear mixed models. Lesaffre and Verbeke [17] studied how much case-weight perturbation impacts

∗Corresponding author. Email: triaswahyuni.rakhmawati@uhasselt.be

parameter estimates; such perturbations refer to infinitesimal deviations from a subject's contribution to the log-likelihood. Their proposal has several attractive features. First, it distinguishes influence in fixed-effects parameters from that in variance components. Second, for each of these parameter subsets, influence is decomposed in interpretable components. Third, the influence diagnostics are computationally inexpensive, once the mixed model is fitted.

The GLMM has received less attention, even though Ouwens, Tan, and Berger [24] applied local influence to count data. An important complication is that the (log-)likelihood function does not admit a closed form. Hence, their derivations were numerical in nature, which makes it less evident to derive meaningful influence components.

Here, we extend local influence for the GLMM in several ways. First, we consider outcomes of binary, count, and time-to event type. Second, using the extension proposed by Molenberghs, Verbeke, and Demétrio [20] and Molenberghs *et al* [21], we flexibly allow for overdispersion in the GLMM, by introducing conjugate random effects, in addition to normal ones. This model is referred to as the combined model. Third, apart from numerical derivations of local influence, we examine two alternative routes: (a) closed forms for the marginal likelihood such as proposed in Molenberghs *et al* [21] and (b) the marginal likelihood with integral form. The closed forms in (a) do not always exist; while they are available for the probit-(beta-)normal, Poisson-(gamma-)normal, and Weibull-(gamma-)normal, they are not for the logit-(beta-)normal. Even when they do, they may be somewhat unwieldy and therefore, route (b) is more promising. Fourth, interpretable components are derived, allowing to get a better perspective on the data-analytic consequences of candidate influential subjects. In other words, once influential subjects have been identified, it can be examined precisely which aspects lead to such influences.

The paper is organized as follows. In Section 2 four case studies are introduced, two counts, one made up of binary, and one of time-to-event type. Their analyses are reported in Section 6. Section 3 describes the generalized model based on the exponential family. Section 4 reviews the essence of local-influence theory. The LMM case is sketched in Section 5.1, and we show that using the integral form of the log-likelihood leads to exactly the same expressions. The Poisson, probit, logit, and Weibull cases are studied in Sections 5.2–5.5.

2. Case Studies

2.1 *A Clinical Trial in Epileptic Patients*

The data considered here are obtained from a randomized, double-blind, parallel group multi-center study for the comparison of placebo with a new anti-epileptic drug (AED), in combination with one or two other AED's. The study is described in Faught *et al* [12]. The randomization of epilepsy patients took place after a 12-week baseline period that served to stabilize the use of AED's, and during which the number of seizures were counted. After that period, 45 patients were assigned to the placebo group, 44 to the active (new) treatment group. Patients were then measured weekly. Patients were followed during 16 weeks, after which they were entered into a long-term open-extension study. Some patients were followed for up to 27 weeks. The outcome of interest is the number of epileptic seizures experienced during the most recent week. The research question is whether or not the additional new treatment reduces the number of epileptic seizures.

2.2 *Headache Study*

This dataset has been reported by McKnight and Van Den Eeden [18]. The experiment has been done using a two-treatment, double-blind crossover design; the number of headaches per week is repeatedly measured during 5 weeks of experiment. The study objective was to investigate whether aspartame causes headaches in subjects who believe they experience aspartame-induced headaches. Twenty-seven volunteers who responded to newspaper advertisements were randomized to one of four treatment regimens. Each regimen began with a seven-day placebo run-in period followed by four treatment periods of seven days each. Each treatment period was separated by a "washout day." Both aspartame (A), given at 30 mg/kg/day, and placebo (P) were administered in capsules of three doses per day. The four possible orderings of treatment after the run-in period were APAP, APPA, PAPA and PAAP. Most of the run-in periods were done within 7 days, yet some of the periods were smaller.

2.3 *A Clinical Trial in Onychomycosis*

These data come from a randomized, double-blind, parallel group, multicenter study for the comparison of two oral treatments (A and B) for toenail dermatophyte onychomycosis (TDO; [10]). TDO is a common toenail infection, difficult to treat, with prevalence exceeding 2% [25]. Anti-fungal compounds, classically used for treatment of TDO, need to be taken until the whole nail has grown out healthily. The development of such new compounds, has reduced the treatment duration to 3 months. The aim of the present study was to compare the efficacy and safety of 12 weeks of continuous therapy with A or B. Twice 189 patients were randomized. Subjects were followed during 3 months of treatment and followed further until month 12. Measurements were taken at 0, 1, 2, 3, 6, 9, and 12 months. The outcome of interest is severity of infection (0: severe; 1: non-severe). The estimand is the difference in slope over time between the arms.

2.4 *Recurrent Muscle Soreness*

These data come from Hosmer and Lemeshow [14]. The study of two treatment modalities was aimed at reducing the occurrence of muscle soreness among 400 middle-aged men in the beginning of weight training. Subjects were randomized over two instructional programs designed to prevent muscles soreness. The control treatment consisted of standard written brochures and instructions used by the health club to explain the proper technique, including the suggestions for frequency and duration of training. The new method included 1 hour with a personal trainer as well as brochures. The subjects were followed for some time and the dates on which muscles soreness limited the prescribed workout were recorded, converted to number of days between soreness episodes. All subjects had between one and four muscle soreness episodes. The start and end of each episode is recorded, together with the status indicator to denote whether the end of the episode corresponds to a muscle soreness or not.

3. Generalized Linear Mixed Models

The generalized linear mixed model [2, 11, 28] is arguably the most frequently used random-effects model in the context of (non-)Gaussian repeated measurements, extending both generalized linear models for univariate outcomes and linear mixed models [26].

Let Y_{ij} be the jth outcome for subject $i = 1, \ldots, N$, $j = 1, \ldots, n_i$ and group the n_i

measurements into vector $\boldsymbol{Y}_i$. Assume that, given q-dimensional random effects

$$\boldsymbol{b_i} \sim N(\boldsymbol{0}, D), \tag{1}$$

the Y_{ij}'s are independent with model

$$\begin{aligned} f_i(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi}, \phi) &= \exp\left\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\right\}, \\ \eta[\psi'(\lambda_{ij})] = \eta(\mu_{ij}) &= \eta[E(Y_{ij}|\boldsymbol{b_i}, \boldsymbol{\xi})] = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i} \end{aligned} \tag{2}$$

for a known link function $\eta(\cdot)$, with $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ p- and q-dimensional vectors of known covariate values, $\boldsymbol{\xi}$ a p-dimensional vector of unknown fixed regression coefficients, and with ϕ a scale (overdispersion) parameter. Let $\phi(\boldsymbol{b_i}|D)$ be the multivariate normal density with mean $\boldsymbol{0}$ and variance D. The marginal likelihood function is:

$$L(\boldsymbol{\vartheta}, D) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{\vartheta}, \boldsymbol{b_i})\ \phi(\boldsymbol{b_i}|D)\ d\boldsymbol{b_i}.$$

Here, $\boldsymbol{\vartheta}$ groups all parameters in the conditional model for $\boldsymbol{Y}_i$ given the random effects. Not always is there a closed form for the integral in (3), nor for the corresponding moments. The most notorious counterexample is the logit-normal model, where (2) uses the logit link. While a suite of computational techniques has been derived to approximate the likelihood numerically, e.g., using Taylor series expansions and numerical integration, it poses further challenges when additional calculations are requested. We are in this position, because local influence starts from the likelihood (see Section 4).

3.1 *The Linear Mixed Model for Gaussian Data*

The hierarchically specified linear mixed-effects model takes the form [26]:

$$\boldsymbol{Y}_i|\boldsymbol{b_i} \sim N(X_i\boldsymbol{\xi} + Z_i\boldsymbol{b_i}, \Sigma_i), \tag{3}$$

where $\boldsymbol{\xi}$ is a vector of fixed effects, and X_i and Z_i are design matrices. The rows of $X_i\boldsymbol{\xi} + Z_i\boldsymbol{b_i}$ are made up by the linear predictors (2). Evidently, $\boldsymbol{b_i}$ is as specified in (1). The corresponding marginal model, needed for maximum likelihood estimation and hence the corner stone for local influence [17] obtains easily and is, again, of a multivariate normal form:

$$\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\xi}, V_i = Z_iDZ_i' + \Sigma_i). \tag{4}$$

3.2 *The Poisson-Normal and Poisson-Gamma-Normal Models for Count Data*

From the general developments above, the Poisson-normal model is:

$$Y_{ij} \sim \text{Poi}(\lambda_{ij}), \tag{5}$$
$$\lambda_{ij} = \exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}\right), \tag{6}$$

and $\boldsymbol{b_i}$ as in (1). Molenberghs, Verbeke, and Demétrio [20] and Molenberghs *et al* [21] derived a closed form for the marginal model:

$$\mathrm{P}(\boldsymbol{Y_i} = \boldsymbol{y_i}) = \frac{1}{\prod_{j=1}^{n_i} y_{ij}!} \sum_{\boldsymbol{t}} \frac{(-1)^{\sum_{j=1}^{n_i} t_j}}{\prod_{j=1}^{n_i} t_j!} \cdot \exp\left[\sum_{j=1}^{n_i} (y_{ij} + t_j)\boldsymbol{x}'_{ij}\boldsymbol{\xi}\right]$$
$$\times \exp\left\{\frac{1}{2}\left[\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}'_{ij}\right] D \left[\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}_{ij}\right]\right\}. \quad (7)$$

The vector-valued index $\boldsymbol{t} = (t_1, \ldots, t_{n_i})$ ranges over all non-negative integer vectors.

When overdispersion is accommodated, as in Molenberghs, Verbeke, and Demétrio [20] and Molenberghs *et al* [21]), (5) changes to

$$Y_{ij} \sim \mathrm{Poi}(\theta_{ij}\lambda_{ij}), \quad (8)$$

with λ_{ij} as in (6) and $\theta_{ij} \sim \mathrm{Gamma}(\alpha_j, \beta_j)$. The joint distribution now is:

$$\mathrm{P}(\boldsymbol{Y_i} = \boldsymbol{y_i}) = \sum_{\boldsymbol{t}} \left[\prod_{j=1}^{n_i} \binom{y_{ij}+t_j}{y_{ij}} \cdot \binom{\alpha_j + y_{ij} + t_j - 1}{\alpha_j - 1} \cdot (-1)^{t_j} \cdot \beta_j^{y_{ij}+t_j}\right]$$
$$\times \exp\left(\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{x}'_{ij}\boldsymbol{\xi}\right)$$
$$\times \exp\left\{\frac{1}{2}\left[\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}'_{ij}\right] D \left[\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}_{ij}\right]\right\}. \quad (9)$$

For identification, write $\beta_j = 1/\alpha_j$. The modeler may choose α_j and β_j terms free of j.

While Zeger, Liang, and Albert [29] derived a closed form for the mean function only, Molenberghs *et al* [21] thus provided closed forms for all of the moments and for the joint marginal distribution; they did so for the combined-model extension of the GLMM, and hence for the GLMM itself. This opens avenues for local influence and corresponding interpretable components, which goes well beyond what was done in the literature thus far (e.g., [24]).

3.3 *The Probit-Normal Model for Binary and Binomial Data*

A probit-normal model is specified by $Y_{ij} \sim Bin(\lambda_{ij}, n_{ij})$ and

$$\lambda_{ij} = \Phi_1(\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}). \quad (10)$$

Molenberghs *et al* [21] showed that the marginal joint distribution is:

$$f_{n_i}(\boldsymbol{y}_i = \boldsymbol{1}) = \Phi_{n_i}(X_i\boldsymbol{\xi}; L_{n_i}^{-1}), \quad (11)$$

with $L_{n_i} = I_{n_i} - Z_i\left(D^{-1} + Z_i'Z_i\right)^{-1} Z_i'$. Of course, this is only the probability of a (so-called success) sequence consisting of ones. All other joint probabilities are derived by the usual combination rules [21]. When overdispersion is allowed for, then again λ_{ij} is multiplied by $\theta_{ij} \sim \text{Beta}(\alpha, \beta)$ and the joint distribution becomes:

$$f_{n_i}(\boldsymbol{y}_i = \boldsymbol{1}) = \left(\frac{\alpha}{\alpha+\beta}\right)^{n_i} \cdot \Phi_{n_i}(X_i\boldsymbol{\xi}; L_{n_i}^{-1}) \tag{12}$$

Should the logit link be used, there is no closed form available. Of course, the approximation rule for the logit by the probit function can be used [16, 21, 29]:

$$f_{n_i}(\boldsymbol{y}_i = \boldsymbol{1}) \approx \Phi_{n_i}\left(cX_i\boldsymbol{\xi}; \widetilde{L}_{n_i}^{-1}\right), \tag{13}$$

with $\widetilde{L}_{n_i} = I_{n_i} - c^2 Z_i \left(D^{-1} + Z_i'Z_i\right)^{-1} Z_i'$ and $c = (16\sqrt{3})/(15\pi)$.

3.4 *The Weibull-Normal Model for Time-to-event Data*

In the Weibull case, the corresponding model is

$$f(\boldsymbol{y}_i|\boldsymbol{\theta}_i, \boldsymbol{b_i}) = \prod_{j=1}^{n_i} \lambda\rho y_{ij}^{\rho-1} e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\boldsymbol{z}'_{ij}\boldsymbol{b_i}} e^{-\lambda y_{ij}^{\rho} e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\boldsymbol{z}'_{ij}\boldsymbol{b_i}}}, \tag{14}$$

with $\boldsymbol{b_i}$ as in (1). The joint distribution is [21]:

$$f(\boldsymbol{y}_i) = \sum_{(t_1,\ldots,t_{n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{t_j}}{t_j!} \lambda^{t_j+1} \rho y_{ij}^{(t_j+1)\rho-1} \times \exp\left\{(t_j+1)\left[\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \frac{1}{2}(t_j+1)\cdot \boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}\right]\right\}. \tag{15}$$

Similar to the Poisson case, $(t_1, \ldots, t_{n_i})$ ranges over all non-negative integer vectors. When overdispersion is allowed for, the Weibull-Gamma-Normal model is:

$$f(\boldsymbol{y}_i|\boldsymbol{\theta}_i, \boldsymbol{b_i}) = \prod_{j=1}^{n_i} \lambda\rho\theta_{ij} y_{ij}^{\rho-1} e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\boldsymbol{z}'_{ij}\boldsymbol{b_i}} e^{-\lambda y_{ij}^{\rho}\theta_{ij} e^{\boldsymbol{x}'_{ij}\boldsymbol{\xi}+\boldsymbol{z}'_{ij}\boldsymbol{b_i}}}, \tag{16}$$

with now also $\theta_{ij} \sim \text{Gamma}(\alpha_j, \beta_j)$, leading to the closed form:

$$f(\boldsymbol{y}_i) = \sum_{(t_1,\ldots,t_{n_i})} \prod_{j=1}^{n_i} \frac{(-1)^{t_j}}{t_j!} \frac{\Gamma(\alpha_j+t_j+1)\beta_j^{t_j+1}}{\Gamma(\alpha_j)} \lambda^{t_j+1} \rho y_{ij}^{(t_j+1)\rho-1} \times \exp\left\{(t_j+1)\left[\boldsymbol{x}'_{ij}\boldsymbol{\xi} + \frac{1}{2}(t_j+1)\cdot \boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}\right]\right\}. \tag{17}$$

4. Review of General Theory for Local Influence

4.1 *Standard Approach*

Local influence was presented by Cook [8] and used by many authors since. The impact of individuals and measurements on the analysis is assessed by comparing standard maximum likelihood estimates with those resulting from slightly perturbing the contribution of an individual or measurement. The method is to be contrasted with global influence (case deletion), where impact is assessed by simply deleting an individual or measurement. While local influence comes with a certain amount of technicality, it is easy and fast to calculate, and in many cases leads to interpretable components of influence. The existence of such interpretable components is often, and also here, a major rationale for using the method. Lesaffre and Verbeke [17] introduced influence assessment for the linear mixed model. A review of several diagnostic procedures for the linear mixed model is given in Mun and Lindstrom [22]. Verbeke *et al* [27] used local influence for longitudinal Gaussian data with dropout, while incomplete binary data were studied by Jansen *et al* [15]. Verbeke and Molenberghs [26] and Molenberghs and Verbeke [19] reviewed the method and provide ample references.

Ouwens, Tan, and Berger [24] applied local influence to the Poisson-normal model. We will follow their steps, but with extensions in three directions. First, we will provide closed-form expressions, based on an analytical form for the marginal likelihood function, as well as based on an integral form for the said likelihood. Second, we consider three important cases: binary, count, and time-to-event. Third, extensions will be constructed to allow for overdispersion in all of these settings. Some authors considered specific extensions as well. For example, Chen, Fu, and Wang [4] considered local influence for zero-inflated Poisson mixtures.

Let the log-likelihood for the generalized linear mixed model or its combined extension take the form

$$\ell(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{N} \ell_i(\boldsymbol{\theta}), \tag{18}$$

in which $\ell_i(\boldsymbol{\theta})$ is the contribution of the ith individual to the log-likelihood. Let

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) \;=\; \sum_{i=1}^{N} \omega_i \ell_i(\boldsymbol{\theta}), \tag{19}$$

now denote the perturbed version of $\ell(\boldsymbol{\theta})$, depending on an N-dimensional vector $\boldsymbol{\omega}$ of weights, assumed to belong to an open subset Ω of $I\!R^N$. The original log-likelihood (18) follows for $\boldsymbol{\omega} = \boldsymbol{\omega_0} = (1, 1, \ldots, 1)'$. Other perturbation schemes are possible [26]. Let $\widehat{\boldsymbol{\theta}}$ be the maximum likelihood estimator for $\boldsymbol{\theta}$, obtained by maximizing $\ell(\boldsymbol{\theta})$, and let $\widehat{\boldsymbol{\theta}}_\omega$ denote the estimator for $\boldsymbol{\theta}$ under $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$. Cook [8] proposed to measure the distance between $\widehat{\boldsymbol{\theta}}_\omega$ and $\widehat{\boldsymbol{\theta}}$ by the likelihood displacement: $\mathrm{LD}(\boldsymbol{\omega}) \;=\; 2\left(\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\right)$. $\mathrm{LD}(\boldsymbol{\omega})$ will be large if $\ell(\boldsymbol{\theta})$ is strongly curved at $\widehat{\boldsymbol{\theta}}$. A graph of $\mathrm{LD}(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ brings out information on the influence of case-weight perturbations. The graph is the geometric surface formed by the values of the $(N+1)$-dimensional vector

$$\boldsymbol{\xi}(\boldsymbol{\omega}) \;=\; \begin{pmatrix} \boldsymbol{\omega} \\ \mathrm{LD}(\boldsymbol{\omega}) \end{pmatrix}$$

as $\boldsymbol{\omega}$ varies throughout Ω. Following Cook [8] and Verbeke and Molenberghs [26], we will refer to $\boldsymbol{\xi}(\boldsymbol{\omega})$ as an influence graph.

Zhu and Lee [30] and Zhu *et al.* [31] proposed another approach to deal with the measurement of the distance between $\widehat{\boldsymbol{\theta}}_\omega$ and $\widehat{\boldsymbol{\theta}}$. Instead of using the observed-data log-likelihood, their method is applied to the objective function that features in the expectation step of the EM algorithm. Because this function is usually denoted by Q, their method is known as Q-displacement. In this paper, we focused on the used of the likelihood displacement LD($\boldsymbol{\omega}$) [8].

Cook [8] derived a convenient computational scheme. Let $\boldsymbol{\Delta}_i$ be the s-dimensional vector of second-order derivatives of $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$, w.r.t. ω_i and all components of $\boldsymbol{\theta}$, and evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Also, write Δ for the $s \times r$ matrix with $\boldsymbol{\Delta}_i$ in the ith column. Let $\ddot{L}$ denote the $s \times s$ matrix of second derivatives of $\ell(\boldsymbol{\theta})$, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. For any unit vector $\boldsymbol{h}$ in Ω, it follows that:

$$C_h = 2\left|\, \boldsymbol{h}'\Delta'\ddot{L}^{-1}\Delta\boldsymbol{h} \,\right|. \tag{20}$$

Various choices for $\boldsymbol{h}$ have received attention. First, as will be done here, one can focus on subject i only, by choosing $\boldsymbol{h} = \boldsymbol{h_i}$, the zero vector with a sole 1 in the ith position. Local influence then is

$$C_i \;\equiv\; C_{h_i} = 2\left|\, \boldsymbol{\Delta}_i'\ddot{L}^{-1}\boldsymbol{\Delta}_i \,\right|. \tag{21}$$

Second, $\boldsymbol{h} = \boldsymbol{h}_{\max}$ can be considered, the direction of maximal normal curvature [26]. Expressions can be derived when only a sub-vector of the parameter vector is of interest. See Supplementary Materials (Section S.1).

4.2 *Proceeding When Faced With a Complicated Likelihood*

As will be reviewed in Section 5.1.1, Lesaffre and Verbeke [17] proceeded by deriving local influence based on the explicit expression of the marginalized linear mixed model. While there are marginal expressions available for the Poisson, probit, and Weibull cases (Sections 3.2–3.4), these are involved. This is why we also proceed in two alternative ways. The first one consists of using integral expression (3), essentially combined with the property that integration and derivation can be interchanged under mild regularity conditions. Importantly, this route still allows for the derivation of interpretable components. A further alternative consists of choosing a fully numerical route, as in Ouwens, Tan, and Berger [24].

5. Local Influence for Generalized Linear Mixed and Combined Models

5.1 *Local Influence for the Linear Mixed Model*

5.1.1 *Standard Approach, Based on the Marginal Likelihood*

The backdrop for our developments is the method as derived for the linear mixed model [26]. They started from the marginal likelihood (4) directly. For this model, this is easy to do and hence a natural choice. We will review their derivations, with details relegated to the Supplementary Appendix Materials (Section S.2). We will then proceed alternatively by an integral-based approach. This will provide the basis for the analogous calculations in the non-Gaussian cases.

For the covariance structure, we assume conditional independence, i.e., $\Sigma_i = \sigma^2 I_{n_i}$, with I_{n_i} the $n_i \times n_i$ identity matrix.

It is advantageous that C_i admits closed form (21). Lesaffre and Verbeke [17] decomposed C_i into five interpretable components. Let $\boldsymbol{\mathcal{R}_i}$, $\mathcal{X}_i$, and $\mathcal{Z}_i$ denote the "standardized" residuals and covariates for the ith individual, defined by $\boldsymbol{\mathcal{R}_i} = V_i^{-1/2}\boldsymbol{r_i}$, $\mathcal{X}_i = V_i^{-1/2}X_i$, and $\mathcal{Z}_i = V_i^{-1/2}Z_i$, respectively, with $\boldsymbol{r_i} = \boldsymbol{y_i} - X_i\widehat{\boldsymbol{\xi}}$. Further, for a matrix A, let $\|A\| = \sqrt{\text{tr}(A'A)}$ be the Frobenius norm of A [13]. The interpretable components in C_i are then

$$\|\mathcal{X}_i\mathcal{X}_i{}'\|, \quad \|\boldsymbol{\mathcal{R}_i}\|, \quad \|\mathcal{Z}_i\mathcal{Z}_i{}'\|, \quad \|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|, \quad \|V_i^{-1}\|. \tag{22}$$

First, $\|\mathcal{X}_i\mathcal{X}_i{}'\|$ measures the "length" of the standardized covariates in the mean structure and $\|\boldsymbol{\mathcal{R}_i}\|$ is an overall measure for how well the observed data for the ith subject are predicted by the mean structure $X_i\boldsymbol{\xi}$. Second, the components $\|\mathcal{Z}_i\mathcal{Z}_i{}'\|$ and $\|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|$ have a similar meaning, but then for the covariance structure. For example, $\|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|$ will be zero only if V_i equals $\boldsymbol{r_i}\boldsymbol{r_i}{}'$. Note that $\boldsymbol{r_i}\boldsymbol{r_i}{}'$ is an estimate for var($\boldsymbol{y_i}$), which only assumes the mean to be correctly modeled as $X_i\boldsymbol{\xi}$. Therefore, $\|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|$ can be interpreted as a residual, capturing how well the covariance structure of the data is modeled by $V_i = Z_iDZ_i' + \sigma^2 I_{n_i}$. Finally, the fifth component $\|V_i^{-1}\|$ will be large if V_i has small eigenvalues, indicating that the ith subject has little variability.

The decomposition of C_i immediately suggests a practical procedure to find an explanation for the influential nature of an individual, i.e., when C_i is large, we examine the diagnostics. Such plots are useful to graphically inspect the individuals in view of their influence. Thus, it is sensible to start with an index plot of C_i. Following this, the index plots of (22) can be examined. A recurrent practical difficulty with diagnostics is to establish a threshold above which an individual is defined as "remarkable." It follows from (21) that

$$\sum_{i=1}^{N} C_i = -2 \text{ tr}\left(\ddot{L}^{-1}\sum_{i=1}^{N}\boldsymbol{\Delta}_i\boldsymbol{\Delta}_i'\right),$$

which converges to $2s$, for N approaching infinity. As with leverage in linear regression [23, pp. 395–396], one could classify an individual for which C_i is larger than twice the average value (larger than $4s/N$, for N large) as influential. However, unlike for the leverage situation, $2s$ is only the approximate sum of the C_i, which will not be accurate if the model is not correctly specified (such that $\ddot{L}^{-1}\sum_{i=1}^{N}\boldsymbol{\Delta}_i\boldsymbol{\Delta}_i'$ does not converge to I_s) or if N is too small for the asymptotic results to be reliable. In such cases, Lesaffre and Verbeke [17] proposed to replace $2s$ by the actual sum; we then call the ith subject influential if C_i is larger than the cutoff value $2\sum_{i=1}^{N} C_i/N$.

Given decomposition result (S.1.1), it is interesting to consider sub-vectors $\boldsymbol{\xi}$ and $\boldsymbol{\alpha}$ of fixed effects and variance components, respectively, with corresponding influences $C_i(\boldsymbol{\xi})$ and $C_i(\boldsymbol{\alpha})$, respectively. Given that the fixed effects and variance components are asymptotically independent, it follows that $C_i \approx C_i(\boldsymbol{\xi}) + C_i(\boldsymbol{\alpha})$. Lesaffre and Verbeke [17] further showed that $C_i(\boldsymbol{\xi})$ can be decomposed using only the first two components $\|\mathcal{X}_i\mathcal{X}_i{}'\|$ and $\|\boldsymbol{\mathcal{R}_i}\|$, while the last three, $\|\mathcal{Z}_i\mathcal{Z}_i{}'\|$, $\|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|$, and $\|V_i^{-1}\|$, feature in the decomposition of $C_i(\boldsymbol{\alpha})$. Asymptotically therefore, influence for the fixed effects and for the variance components can be scrutinized by studying the first two and the last three interpretable components, respectively.

5.1.2 Integral-based Expression

As mentioned in Section 4, the integral-based approach can be used as an alternative way to alleviate complexities with the explicit marginal likelihood expressions. To prepare for developments of Poisson, probit, logit, and Weibull cases, the calculations have been done first for the linear mixed model setting. Details are in Supplementary Section S.2.2. This integral-based result is identical to the standard one of Lesaffre and Verbeke [17], reported in the previous section. Evidently, the same interpretable components as in (22) ensue.

5.1.3 Fully Numerical Route

The third and final method examined proceeds fully numerically. Observe that (20) is based on the first- and second-order derivatives of the log-likelihood function. Methodologically, a fully numerical derivation is based on replacing derivatives by appropriately precise finite differences of the first and second order, for the score vector and Hessian matrix, respectively. Conveniently, such calculations are routinely done in statistical software packages as part of the log-likelihood maximization process. All that is needed is extracting this information from the package. For the score, individual subjects' contributions are needed, as is clear in particular from (21). The advantage of this approach is straightforward implementation for the models considered here but also for other models with perturbation scheme (19) for the log-likelihood, provided that the score and Hessian functions are numerically available.

Even though jointly considering the numerical approach and the explicit route appears redundant, it is beneficial to make use of both. We referred to the computational ease of the numerical method. At the same time, the explicit calculations can be used to also calculate the influence components, for enhanced interpretation. This route is followed, using the SAS procedure NLMIXED.

5.2 Local Influence for the Poisson-normal Model

In this section, local influence for the Poisson-normal model is studied. In the Supplementary Materials (Section S.3.1), it is shown that, while one could set out from the explicit marginal distribution, the infinite sum that it contains inhibits both convenient expressions and interpretable components. We therefore prefer an integral-based approach, the details of which are given in the Supplementary Materials (Section S.3.2). Writing

$$I_i = \int \exp\left\{\widetilde{f}(\boldsymbol{y_i}) + \widetilde{f}(\boldsymbol{b_i})\right\} d\boldsymbol{b_i},$$

$$\widetilde{f}(\boldsymbol{y_i}) = \sum_{j=1}^{n_i} \left\{c_{ij}y_{ij} - \exp(\gamma_{ij} + c_{ij})\right\},$$

$$\widetilde{f}(\boldsymbol{b_i}) = -\frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i},$$

$$A_i = \exp\left\{\widetilde{f}(\boldsymbol{y_i}) + \widetilde{f}(\boldsymbol{b_i})\right\},$$

it follows that

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n_i} \{y_{ij} - E(y_{ij}|\boldsymbol{b_i})\} \boldsymbol{x}_{ij} = \sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij},$$

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial d_{jk}} = -\frac{1}{2}(2 - \delta_{jk}) \left\{(D^{-1})_{jk} - (D^{-1}D^{-1})_{jk}\text{Var}(\boldsymbol{b_i})\right\},$$

where d_{jk} is a component of D and δ_{jk} is one if j is equal to k, and zero otherwise. Also, by $\text{Var}(\boldsymbol{b_i})$ we mean $\sum_{k=1}^{q} \text{Var}(b_{ik})$.

Interpretable expressions can now be derived. To this end, in the Supplementary Materials (Section S.3.2), we first show that

$$||\boldsymbol{\Delta}_i||^2 = \left(\sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}\right)\left(\sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}\right)' + \sum_{k,l}\left\{-\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl}\text{Var}(\boldsymbol{b_i})\right\}^2.$$

Let $C_i = C_{1i} + C_{2i}$ with:

$$C_{1i} = 2||\ddot{L}^{-1}||\ ||\boldsymbol{r_i}\boldsymbol{x}_i||^2 \cos(\varphi_i), \tag{23}$$

$$C_{2i} = \frac{1}{2}||\ddot{L}^{-1}||\ ||(D^{-1})_{kl} - (D^{-1}D^{-1})_{kl}\text{Var}(\boldsymbol{b_i})||^2 \cos(\varphi_i), \tag{24}$$

where $\boldsymbol{r_i}\boldsymbol{x}_i = \sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}$. Note that C_{1i} and C_{2i} are the contributions of subject i to local influence C_i from $\boldsymbol{\beta}$ and D, respectively. Now, C_{1i} and C_{2i} can be shown to equal:

$$C_{1i} = 2||\ddot{L}^{-1}||\ ||\boldsymbol{x}_i\boldsymbol{x}_i'||\ ||\boldsymbol{r_i}||^2 \cos(\alpha_i)\cos(\varphi_i), \tag{25}$$

$$\begin{aligned} C_{2i} = \frac{1}{2}||\ddot{L}^{-1}||\cos(\varphi_i) \times \Big[&\text{tr}\left\{(D^{-1})^2_{kl}\right\} - \text{tr}\left\{2(D^{-1})_{kl}(D^{-1}D^{-1})_{kl}\text{Var}(\boldsymbol{b_i})\right\} \\ &+\text{tr}\left\{(D^{-1}D^{-1})^2_{kl}\text{Var}(\boldsymbol{b_i})^2\right\}\Big], \end{aligned} \tag{26}$$

where $\cos(\alpha_i)$ is the angle between $\text{vec}(\boldsymbol{x}_i\boldsymbol{x}_i')$ and $\text{vec}(\boldsymbol{r_i}\boldsymbol{r_i}')$, and φ_i is the angle between $\text{vec}(-\ddot{L}^{-1})$ and $\text{vec}(\boldsymbol{\Delta}_i\boldsymbol{\Delta}_i')$. Hence, the interpretable components of C_i in the case of the Poisson-normal model can be described using the 'length of the fixed effect' ($||\boldsymbol{x}_i\boldsymbol{x}_i'||$), the 'squared length of the residual' ($||\boldsymbol{r_i}||^2$), and the 'squared of random effect variability' ($\text{Var}(\boldsymbol{b_i})^2$).

5.3 *Local Influence for the Probit-normal Model*

Given the numerical approach of Section 5.1.3, we will focus on the explicit calculations, using only the integral method. Derivations are in the Supplementary Materials (Section S.4). The binomial probability, conditional on the random effects, is:

$$\text{P}(\boldsymbol{y}_i|\boldsymbol{\xi}, \boldsymbol{b_i}) = \prod_{j=1}^{n_i} \lambda_{ij}^{y_{ij}}(1 - \lambda_{ij})^{(1-y_{ij})}, \tag{27}$$

where λ_{ij} is defined by (10). The joint marginal probability of success is:

$$f(\boldsymbol{y}_i = 1) = \frac{1}{(2\pi)^{q/2}|D|^{1/2}} \int \left(\prod_{j=1}^{n_i} \Phi_{n_i}(\boldsymbol{X}_i'\boldsymbol{\xi} + \boldsymbol{Z}_i'\boldsymbol{b_i}) \right) \exp(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i})d\boldsymbol{b_i}. \quad (28)$$

The first derivatives are:

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial \boldsymbol{\xi}} = [I - (\boldsymbol{X}_i\boldsymbol{\beta})^{-1}]\boldsymbol{X}_i,$$

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial d_{jk}} = \frac{3}{2}L^{-1}\left(I_{n_i} - Z_iM_iM_i'(D^{-1}D^{-1})_{jk}Z_i'\right),$$

where $M_i = \left(D^{-1} + Z_i'Z_i\right)^{-1}$. It also follows that

$$||\boldsymbol{\Delta}_i||^2 = [I - (\boldsymbol{X}_i\boldsymbol{\beta})^{-1}]^2\boldsymbol{X}_i\boldsymbol{X}_i' + \sum_{k,l}\frac{9}{4L^2}\left(I_{n_i} - Z_iM_iM_i'(D^{-1}D^{-1})_{jk}Z_i'\right)^2.$$

Thus, also for this case, the components $||\boldsymbol{X_i}||^2$ and $||\boldsymbol{Z_i}\boldsymbol{Z_i}'||^2$ turn up.

5.4 *Local Influence for the Logit-normal Model*

The derivations for the logit-normal case are given in the Supplementary Materials (Section S.5).

Evidently, the same binomial expression (27) is used, but now with $\text{logit}(\lambda_{ij}) = \boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b_i}$. The marginal joint density function is:

$$f(\boldsymbol{y}_i = 1) = \frac{1}{(2\pi)^{q/2}|D|^{1/2}} \int \prod_{j=1}^{n_i} \lambda_{ij} \exp\left(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}\right) d\boldsymbol{b_i}.$$

The derivatives take the form:

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \int \frac{1}{1+\exp(\mu_{ij})}\tilde{\tau}(\boldsymbol{b_i}|\boldsymbol{y}_i)d\boldsymbol{b_i},$$

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial d_{jk}} = -\frac{1}{2}(2-\delta_{jk})\left\{(D^{-1})_{jk} - (D^{-1}D^{-1})_{jk}\text{Var}(\boldsymbol{b_i})\right\},$$

where $\mu_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b_i}$. It also follows that

$$||\boldsymbol{\Delta}_i||^2 \propto \left\{\left(\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}\right)\left(\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}\right)' + \sum_{k,l}\left(-\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl}\text{Var}(\boldsymbol{b_i})\right\}^2.$$

Reconstructing the fixed- and random-effects components, respectively, like in the Poisson case, leads to $C_{1i} = 2||\ddot{L}^{-1}||\ ||\boldsymbol{x}_i||^2\cos(\varphi_i)$ and C_{2i} as in (26). Hence, the interpretable components of C_i for the logit-normal model can be described using the length of fixed effect ($||\boldsymbol{x}_i||^2$) and the squared random-effects variability, $\text{Var}(\boldsymbol{b_i})^2$ (i.e., the sum of all variances), in analogy with the Poisson-normal model. The same is true for the Weibull-normal model, as will be seen next.

5.5 *Local Influence for the Weibull-normal Model*

The general Weibull model is given by (14). By means of the derivations given in the Supplementary Materials (Section S.6), the derivatives take the form:

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} - \lambda \sum_{j=1}^{n_i} y_{ij}^{\rho} \boldsymbol{x}_{ij} \exp(\boldsymbol{\mu_{ij}}),$$

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial d_{jk}} = -\frac{1}{2}(2 - \delta_{jk}) \left[(D^{-1})_{jk} - (D^{-1}D^{-1})_{jk} \mathrm{Var}(\boldsymbol{b_i}) \right],$$

where $\delta_{jk} = 1$ if $j = k$ and 0 otherwise. It further follows that

$$||\boldsymbol{\Delta}_i||^2 = \left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \right) \left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \right)' - 2 \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \boldsymbol{Q}_i' + \boldsymbol{Q}_i \boldsymbol{Q}_i'$$

$$+ \sum_{k,l} \left\{ -\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl} \mathrm{Var}(\boldsymbol{b_i}) \right\}^2,$$

where $\boldsymbol{Q}_i = \lambda \sum_{j=1}^{n_i} y_{ij}^{\rho} \boldsymbol{x}_{ij} \exp(\boldsymbol{\mu_{ij}})$. Like in the Poisson-normal and binary-normal cases, a decomposition $C_i = C_{1i} + C_{2i}$ follows, with $C_{1i} = 2||\ddot{L}^{-1}|| \ \left\{ ||\boldsymbol{x}_i||^2 - 2\boldsymbol{x}_i \boldsymbol{Q}_i + ||\boldsymbol{Q}_i||^2 \right\} \cos(\varphi_i)$ and C_{2i} as in (26). Hence, interpretable components analogous to the earlier settings arise.

6. Analysis of Case Studies

6.1 *A Clinical Trial in Epileptic Patients*

We start from the Poisson-normal (P-N) and Poisson-gamma-normal (PGN) models formulated by Molenberghs, Verbeke, and Demétrio [20] and Molenberghs *et al* [21], with Poisson parameter:

$$\ln(\lambda_{ij}) = \begin{cases} (\xi_{00} + b_i) + \xi_{01} t_j & \text{if placebo} \\ (\xi_{10} + b_i) + \xi_{11} t_j & \text{if treated,} \end{cases} \tag{29}$$

where Y_{ij} represent the number of epileptic seizures patient i experienced during week j, t_j is the time point at which Y_{ij} was measured, and with random intercept $b_i \sim N(0, d)$. Parameter estimates are given in Table 1. Index plots (versus patient ID) for various local influence analyses are given in Figure 2. The top row of the plot represents the total local influence, with subsequent rows representing influence for sub-vectors: fixed effects, random-intercept variance d, and, for the (PGN), the overdispersion parameter α, respectively. Patients #38, #49, and #62 stand out with large total influence C_i when compared to other patients. Importantly, influences show a major drop when switching from (P-N) to (PGN). This is most prominently seen for #38. For an explanation, turn to the right hand panel of Figure 1. Patient #38 (and to some extent also #62 on the left hand side) alternates periodically between very high numbers of episodes and periods virtually without. This implies that their mean, variance, and association structure are rather different from the majority of subjects. The impact on the mean structure, by way of the fixed effects, is evident in the second row. For the (P-N) it is less clear when turning to d, but we gain a lot of insight from the (PGN) results. Overall influence and influence on $\boldsymbol{\xi}$ reduce drastically, but there now is clear influence on d and α. What it means is that with these subjects present, the overdispersion parameter helps capturing their anomalous behavior, which 'deflates' d. In other words, adding overdispersion protects the inferentially crucial fixed-effects parameter vector. When removing these subjects, and also #49, little or no influence is left.

Note that the (PGN) model fitted to the full dataset exhibits a smaller value for α, which corresponds to more overdispersion (no overdispersion corresponds to α approaching $+\infty$), while it does not vanish with removal of the three subjects. Thus, there appears to be genuine overdispersion in the data, further inflated by the influential subjects.

In agreement with Molenberghs, Verbeke, and Demétrio [20] and Molenberghs *et al* [21], we consider the treatment effect in additive ($\xi_{11} - \xi_{01}$) and multiplicative (ξ_{11}/ξ_{01}) form. Important differences are seen on the additive scale. (P-N) shows no significance ($p = 0.7106$), which is sustained for (PGN), with $p = 0.2225$. Removing the influential subjects leads to a highly significant result for (P-N), with $p = 0.0009$, which changes to the still significant $p = 0.0350$ for (PGN). Hence, the influential subjects mask a treatment effect. This is logical, because the influential subjects exhibit an oscillating behavior, introducing an important source of variability. At the multiplicative level, where the null hypothesis is for the ratio to be 1, the story is nicely confirmed, with $p = 0.6872$ and $p = 0.1166$ for (P-N) and (PGN), respectively; the counterparts after deletion are $p < 0.0001$ and $p = 0.0040$, respectively.

To get further insight as to why these subject have higher influence than others, plots with interpretable components are given in Figure 3: 'squared length of the fixed effects' $||\boldsymbol{x}_i\boldsymbol{x}_i'||$, 'squared length of the residual' $||\boldsymbol{r_i}||^2$, and 'random-effect variability' $\text{Var}(b_i)^2$. It is hardly surprising that #38 stands out in terms of $||\boldsymbol{r_i}||^2$. Influences on #49 and #62 are less pronounced.

Our analysis has provided insight not available from earlier analysis. The influential subjects exhibit a cyclic behavior not observed in the majority of patients, but at the same time well documented. Based on these findings, a focused clinical discussion can take place, to determine the course of action. Options include removal, retention, or even setting up a dedicated study to further scrutinize this sub-population. In this case, a small group of patients with oscillating behavior between two poles has been identified.

6.2 *Headache Study*

For these data, the model of Ouwens, Tan, and Berger [24] is used again:

$$E(Y_{ij}|\boldsymbol{\xi}, b_i) = t_{ij} \exp(\xi_0 + T_{ij}\xi_1 + b_i), \tag{30}$$

where T_{ij} indicates whether either placebo or Aspartame is given to patient i at occasion j, $b_i \sim N(0, d)$, and t_{ij} is the length of this period in days. We consider (P-N) and (PGN).

Figure S.4 in Supplementary Material shows the individual profiles. From this figure, and from some influence graphs in Figure S.5, patients #13, #25, #4, and #10 deserve further investigation. The individual profiles show that the former two have more headaches than is typically the case; the latter two have none. Subjects #4 and #10 show up in the total influence and that on d for the (P-N), while #13 is influential for the fixed effects. No further influences are seen after removal of these four patients. Also, the relatively strong influences in (P-N) essentially disappear when turning to the (PGN). In other words, these influential subjects induce overdispersion which, when accommodated, strongly alleviates their influential status. Parameter estimates (standard errors) are reported in Table 1. While there is a borderline significant treatment effect in the (P-N) fitted to all data ($p = 0.0463$), and a borderline non-significant one in the (PGN) fitted to the full data ($p = 0.0639$), it disappears after removal ($p = 0.2542$ for (P-N) and $p = 0.2962$ for (PGN)). This underscores that a few subjects might drive the alleged treatment effect. Note that the effect of removal in terms of significance is opposite to that in the epilepsy study. The interpretable components do not lead to additional insight (Figures S.6 in Supplementary Material).

6.3 *A Clinical Trial in Onychomycosis*

Molenberghs *et al* [21] assumed $Y_{ij}|b_i \sim \text{Bernoulli}(\pi_{ij})$, where Y_{ij} is severity of infection (1 for severe, 0 for non-severe) for patient i at occasion j, T_i is the treatment indicator (1 for experimental, 0 for standard) for subject, t_j is the time point (months) at which the jth measurement has been taken, and $b_i \sim N(0, d)$. The conditional success probability is expressed as:

$$\text{logit}(\pi_{ij}) = \xi_1(1 - T_i) + \xi_2(1 - T_i)t_{ij} + \xi_3 T_i + \xi_4 T_i t_{ij} + b_i.$$

Both the logit-normal (L-N) and logit-beta-normal (LBN) are fitted. Parameter estimates (standard errors) are displayed in Table 1, with local influence plots in Figure S.7 (in Supplementary Material). Subjects #6, #30, and #53 are detected as influential, overall, and with respect to the fixed effects, in the (L-N). Accommodating overdispersion, hence turning to the (LBN), deflates the magnitude of influence. Likewise, influence is drastically diminished by removing these three subjects. Thus, in case the influential subjects should remain in the analysis, the (LBN) may be the most sensible route forward. Alternatively, in case they are considered anomalous, one can remove them. To decide on which scenario is preferred in this case, we note that all three subjects are unusual: they set out with a sequence of non-severe ratings, but then switch to a severe rating ('0000111' for #6, '0000011' for #30, and '0000001' for #53). Arguably, there is no reason to remove these subjects from analysis, partly also to safeguard randomization. However, it is uncommon to switch from non-severe to severe in this particular way, so these patients must be further clinically scrutinized. Also for these data, the interpretable components do not lead to further insight (Figure S.8 in Supplementary Materials).

The (L-N) and (LBN) lead to borderline significance when applied to the full data [$p = 0.0268$ additively and $p = 0.0560$ multiplicatively for (L-N); $p = 0.0627$ additively and $p = 0.0964$ multiplicatively for (LBN)]. When influential subjects are removed, these values all become highly significant [in the same order, $p < 0.0001$, $p = 0.0007$, $p = 0.0011$, and $p = 0.0099$]. These findings are qualitatevely similar to the epilepsy cases, but different from the headache study.

6.4 *Recurrent Muscle Soreness*

The Weibull-normal (W-N) and Weibull-gamma-normal (WGN) models are considered, with scale parameter $\lambda = 1$, and linear predictor $\eta_{ij} = \xi_0 + b_i + \xi_1 T_i$, where T_i is an indicator for treatment and $b_i \sim N(0, d)$. Parameter estimates (standard errors) are in Table 1. Local influence plots and interpretable components are displayed in Figures S.9 and S.10 in Supplementary Materials, respectively. Unlike in the three previous studies, no subjects stand out. It is clear though, that influence goes down when turning from the (W-N) to the (WGN). It is equally important to see no influence is detected when there happens to be none.

7. Simulation Study

In order to evaluate the relative performance of local influence in the context of the two different models, standard and combined, a small-scale simulation study was conducted. To define realistic simulation scenarios, the various standard models were considered, for the epilepsy data, the onychomycosis data, and the recurrent muscle soreness data, respectively (see Table 1). These parameter estimates were then used as true values in the simulations and plugged into the corresponding model for each of the three data types. This model is then used to generate the response variables Y_{ij}^{new}. Various sets of covariate values from the original dataset were considered; they were kept fixed across simulation runs. Some predetermined influential subjects are chosen prior to the simulation study for each situation. We consider three types of predetermined influential subjects: high, medium, and low. The local influence analysis is run for each simulated dataset. A cut-off value for local influence is defined as $2\sum_{i=1}^{N} C_i/N$ [17]. Every time, 200 replicated datasets were generated.

The simulation results are presented in Tables 2–4. Table 2 shows the summary statistics. While convergence was unproblematic in the count and time-to-event cases, more difficulties were encountered in the binary case: the combined model gave valid result only for 145 simulations. It is known that identifying overdispersion on top of data correlation in the binary case is harder. From this table, it can be seen that the combined model for all data types identified the influential subjects more frequently than the standard GLMM. Observe that the mean of the local influence values for the combined models are lower than those from the GLMM (Table 3). These findings are in line with the analysis from the original datasets, showing that the local influence for combined models are lower than those from the GLMM.

A classification of the predetermined influential subjects is given in Table 4. Most of the highly influential subjects are classified as influential, for both models and all three data types. In contrast, medium- and low-influence subjects are not always recognized.

8. Concluding Remarks

Local influence was studied before as a means to detect outlying subjects, and features thereof, for the linear mixed model and some generalized linear mixed models. We have extended this work in several ways. First, local influence measures are derived for several GLMM: Poisson-normal, logit-normal, probit-normal, and Weibull-normal. Second, also for the extensions of these model that capture overdispersion, i.e., the combined model, influence measures are derived. Third, using the integral form of the log-likelihood, it has been possible to derive interpretable components of influence, like for the LMM, but unlike in earlier influence work for the GLMM. Beyond identifying influential subjects, this allows us to scrutinize which aspects leads to influence on important model parameters and conclusions based there upon.

In all four case studies analyzed, it is seen that accounting for overdispersion alleviates influence, whether for a few outlying subjects or for the dataset as a whole. When there are outlying subjects in the GLMM, it is often seen that removing them leads to reductions similar to switching to the combined model. Of course, these actions are very different and depend on whether one wants to either homogenize the original data or, conversely, retain these subjects for analysis, but then change the model to one that allows for this without undue influence. The combined model is a good candidate for this. This is underscored by the fact that treatment effect assessment can change in different ways upon removing influential subjects. In the epilepsy and onychomycosis studies, treatment effect turns from non- to (highly) significant; in the headache study, a borderline significant effect disappears after removing influential subjects.

Evidently, beyond the distributions considered here, others could be studied as well. For example, with time-to-event data, it is not uncommon to use log-normal rather than Weibull distributions. Our method is generic and has been applied to a collection of distributions; similar calculations would lead to expressions for alternative distributions.

Web Appendices S.1–S.6, referenced in Section 5, are available in conjunction with this paper.

The methodology developed here has been implemented in the SAS software system. Fitting the models is done using the SAS procedure NLMIXED and macros have been developed for the local influence calculations. The codes are available in the Supplementary Materials.

Acknowledgments

Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

References

[1] Beckman, R.J., Nachtsheim, C.J., and Cook, R.D. (1987) Diagnostics for mixed-model analysis of variance. *Technometrics*, **29**, 413–426.
[2] Breslow, N.E. and Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
[3] Chatterjee, S. and Hadi, A.S. (1988) *Sensitivity Analysis in Linear Regression*. New York: John Wiley & Sons.
[4] Chen, X.-D., Fu, Y.-Z., and Wang, X.-R. (2013) Local influence measure of zero-inflated generalized Poisson mixture regression models. *Statistics in Medicine*, **32**, 1294–1312.
[5] Cook, R.D. (1977a) Detection of influential observations in linear regression. *Technometrics*, **19**, 15–18.
[6] Cook, R.D. (1977b) Letter to the editor. *Technometrics*, **19**, 348.
[7] Cook, R.D. (1979) Influential observations in linear regression. *Journal of the American Statistical Association*, **74**, 169–174.
[8] Cook, R.D. (1986) Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, **48**, 133–169.
[9] Cook, R.D. and Weisberg, S. (1982) *Residuals and Influence in Regression*. London: Chapman & Hall.
[10] De Backer, M., De Keyser, P., De Vroey, C., and Lesaffre, E. (1996) A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs. itraconazole 200mg/day–a double-blind comparative trial. *British Journal of Dermatology* **134**, 16–17.
[11] Engel, B. and Keen, A. (1994) A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica*, **48**, 1–22.
[12] Faught, E., Wilder, B.J., Ramsay, R.E., Reife, R.A., Kramer, L.D., Pledger, G.W., and Karim, R.M. (1996) Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages, *Neurology*, **46**, 1684–1690.
[13] Golub, G.H. and Van Loan, C.F. (1989) *Matrix Computations*. (2nd ed.). Baltimore: The Johns Hopkins University Press.

[14] Hosmer, D.W. and Lemeshow, S. (1999) *Applied Survival Analysis: Regression Modelling of Time to Event Data*. Chichester: John Wiley & Sons.
[15] Jansen, I., Molenberghs, G., Aerts, M., Thijs, H., and Van Steen, K. (2003) A Local influence approach applied to binary data from a psychiatric study. *Biometrics*, **59**, 410–419.
[16] Johnson, N.L. and Kotz, S. (1970) *Distributions in Statistics, Continuous Univariate Distributions, Vol. 2.* Boston: Houghton-Mifflin.
[17] Lesaffre, E. and Verbeke, G. (1998) Local influence in linear mixed models. *Biometrics*, **54**, 570–582.
[18] McKnight, B. and Van Den Eeden, S.K. (1993) A conditional analysis for two-treatment multiple period crossover designs with binomial or Poisson outcomes and subjects who drop out. *Statistics in Medicine*, **12**, 825–834.
[19] Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data.* New York: Springer.
[20] Molenberghs, G., Verbeke, G., and Demétrio, C. (2007) An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, 513–531.
[21] Molenberghs, G., Verbeke, G., Demétrio, C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.
[22] Mun, J. and Lindstrom, M.J. (2013) Diagnostics for repeated measurements in linear mixed effects models. *Statistics in Medicine*, **32**, 1361–1375.
[23] Neter, J., Wasserman, W., and Kutner, M.H. (1990) *Applied Linear Statistical Models. Regression, Analysis of Variance and Experimental Designs* (3rd ed.). Homewood, IL: Richard D. Irwin, Inc.
[24] Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2001). Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, **57**, 1166–1172.
[25] Roberts, D.T. (1992) Prevalence of dermatophyte onychomycosis in the United Kingdom: results of an omnibus survey. *British Journal of Dermatology*, **126**, 23–27.
[26] Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data.* New York: Springer-Verlag.
[27] Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M.G. (2001) Sensitivity analysis for non-random dropout: a local influence approach. *Biometrics*, **57**, 7–14.
[28] Wolfinger, R. and O'Connell, M. (1993) Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, **48**, 233–243.
[29] Zeger, S.C., Liang, K.-Y., and Albert, P.S. (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049–1060.
[30] Zhu, H. and Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **63**, 111–126.
[31] Zhu, H., Lee, S., Wei, B.-C., and Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika*, **88**, 727–737.

Table 1. Parameter estimates (standard errors) for the generalized linear mixed and combined models.

Epilepsy		Poisson-normal		Poisson-gamma-normal	
Effect	Par.	Full	#(38,49,62)	Full	#(38,49,62)
Interc. plac.	ξ_{00}	0.818(0.168)	0.903(0.157)	0.911(0.176)	0.907(0.163)
Slope plac.	ξ_{01}	-0.014(0.004)	-0.031(0.005)	-0.025(0.008)	-0.031(0.008)
Interc. treat.	ξ_{10}	0.648(0.170)	0.492(0.162)	0.656(0.178)	0.510(0.169)
Slope treat.	ξ_{11}	-0.012(0.004)	-0.007(0.005)	-0.012(0.007)	-0.009(0.007)
Treat. eff.	$\xi_{11} - \xi_{10}$	0.002(0.006)	0.024(0.007)	0.013(0.011)	0.022(0.011)
Treat. eff.	ξ_{11}/ξ_{10}	0.840(0.398)	0.236(0.170)	0.475(0.335)	0.281(0.250)
Std. rand. int.	σ	1.076(0.086)	0.982(0.081)	1.063(0.087)	0.969(0.082)
Overdisp. par.	α			2.464(0.211)	3.109(0.329)
Headache		**Poisson-normal**		**Poisson-gamma-normal**	
Effect	Par.	Full	#(4,10,13,25)	Full	#(4,10,13,25)
Intercept	ξ_0	-1.715(0.172)	-1.609(0.136)	-1.710(0.174)	-1.599(0.139)
Treatment	ξ_1	0.283(0.142)	0.187(0.164)	0.289(0.156)	0.187(0.179)
Std. rand. int.	σ	0.695(0.140)	-0.388(0.120)	0.682(0.144)	-0.349(0.137)
Overdisp. par.	α			12.47(16.53)	8.916(9.982)
Onychomycosis		**Logit-normal**		**Logit-beta-normal**	
Effect	Par.	Full	#(6,30,53)	Full	#(6,30,53)
Interc. plac.	ξ_0	-1.630(0.435)	-1.940(0.523)	-1.604(4.026)	-2.420(3.089)
Slope plac.	ξ_1	-0.404(0.046)	-0.430(0.049)	-6.478(1.439)	-6.075(1.264)
Interc. treat.	ξ_2	-1.749(0.448)	-1.604(0.536)	-16.21(3.58)	-15.21(3.02)
Slope treat.	ξ_3	-0.563(0.060)	-0.872(0.100)	-8.075(1.600)	-8.755(1.437
Treat. eff.	$\xi_{11} - \xi_{10}$	-0.159(0.072)	-0.442(0.105)	-1.596(0.858)	-2.680(0.822)
Treat. eff.	ξ_{11}/ξ_{10}	1.394(0.206)	2.028(0.302)	1.246(0.148)	1.441(0.171)
Std. rand. int.	σ	4.015(0.381)	4.814(0.490)	60.88(14.22)	56.47(11.69)
Overdisp. par.	α/β			0.281(0.035)	0.231(0.031)
Muscle Soreness		**Weibull-normal**		**Weibull-gamma-normal**	
Effect	Par.	Full		Full	
Intercept	ξ_0	-3.664(0.1103)		-3.870(0.141)	
Slope	ξ_1	0.352(0.064)		0.404(0.073)	
Shape par.	ρ	1.027(0.027)		1.118(0.045)	
Std. rand. int.	σ	0.242(0.066)		0.199(0.096)	
Overdisp par.	α			5.781(2.174)	

Table 2. Simulation study. The mean for total number of influence subjects across all simulations.

Source	Mean	Std. Dev.	Mean	Std. Dev.
Epilepsy	Poisson-normal		Poisson-gamma-normal	
Total Local Influence (C_i)	6.630	1.122	8.180	0.890
Local Influence(ξ)	6.630	1.067	7.250	0.928
Local Influence (d)	9.220	1.371	8.885	1.048
Local Influence (α)			6.035	1.009
Onychomycosis	Logit-normal		Logit-beta-normal	
Total Local Influence (C_i)	23.920	4.820	33.538	7.947
Local Influence(ξ)	29.790	5.552	27.379	8.856
Local Influence (d)	20.850	3.251	25.841	4.739
Local Influence (α)			43.179	19.089
Muscle Soreness	Weibull-normal		Weibull-gamma-normal	
Total Local Influence (C_i)	21.195	3.350	39.610	4.476
Local Influence(ξ)	2.205	0.494	55.300	4.098
Local Influence(ρ)	40.28	4.152	42.165	3.764
Local Influence (d)	29.985	3.710	25.825	3.698
Local Influence (α)			52.955	80.190

Table 3. Simulation study. The mean of local influence across all simulations.

Source	Mean	Std. Dev.	Mean	Std. Dev.
Epilepsy	Poisson-normal		Poisson-gamma-normal	
Total Local Influence (C_i)	0.663	1.052	0.219	0.011
Local Influence(ξ)	0.622	1.001	0.135	0.008
Local Influence (d)	0.024	0.007	0.021	0.003
Local Influence (α)			0.070	0.007
Onychomycosis	Logit-normal		Logit-beta-normal	
Total Local Influence (C_i)	0.048	0.002	0.042	0.152
Local Influence(ξ)	0.040	0.002	0.020	0.137
Local Influence (d)	0.006	0.001	0.002	0.001
Local Influence (α)			0.014	0.037
Muscle Soreness	Weibull-normal		Weibull-gamma-normal	
Total Local Influence (C_i)	1.446	0.091	0.023	0.001
Local Influence(ξ)	0.090	0.006	0.010	0.00001
Local Influence(ρ)	0.034	0.002	0.005	0.00015
Local Influence (d)	0.940	0.050	0.006	0.001
Local Influence (α)			0.003	0.009

Table 4. Simulation study. Classification of predetermined influential subjects across all simulations.

Category	Subject	Not Influential	Influential	Not Influential	Influential
Epilepsy		Poisson-normal		Poisson-gamma-normal	
High	#5	0	200	0	200
	#38	1	199	0	200
	#49	0	200	0	200
	#62	0	200	0	200
Medium	#2	174	26	1	199
	#16	200	0	200	0
	#60	173	27	151	49
	#73	0	200	0	200
Low	#11	200	0	145	55
	#39	200	0	200	0
	#63	23	177	12	188
	#67	200	0	200	0
Onychomycosis		Logit-normal		Logit-beta-normal	
High	#6	0	200	1	139
	#30	0	200	1	144
	#53	0	200	1	144
	#198	0	200	2	143
Medium	#3	0	200	1	143
	#13	0	200	2	142
	#276	0	200	2	139
	#279	0	200	1	143
Low	#244	94	106	139	1
	#257	94	106	139	1
	#272	152	48	136	1
	#290	152	48	136	1
Muscle Soreness		Weibull-normal		Weibull-gamma-normal	
High	#62	7	193	0	200
	#169	27	173	0	200
	#328	0	200	0	200
	#378	0	200	0	200
Medium	#31	200	0	0	200
	#64	200	0	0	200
	#259	0	200	0	200
	#317	0	200	0	200
Low	#30	200	0	198	2
	#161	200	0	175	25
	#237	0	200	0	200
	#299	0	200	0	200

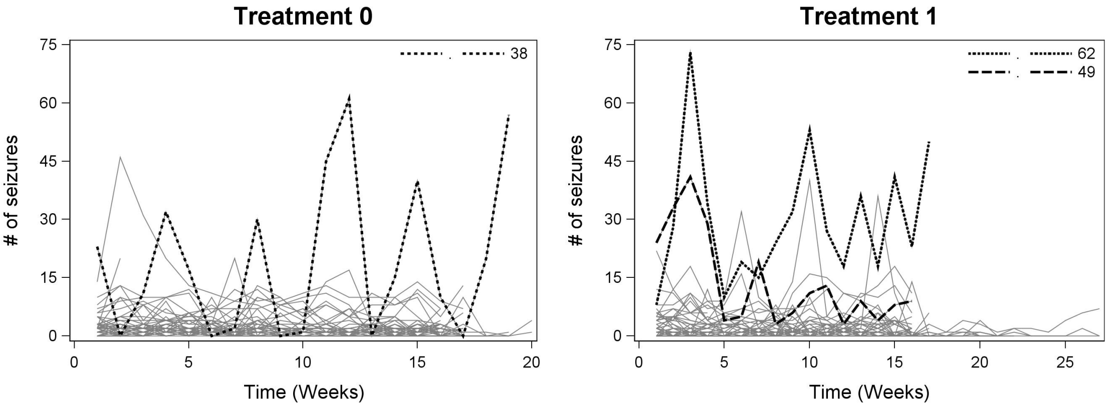

Figure 1. Epilepsy Data. Individual profiles.

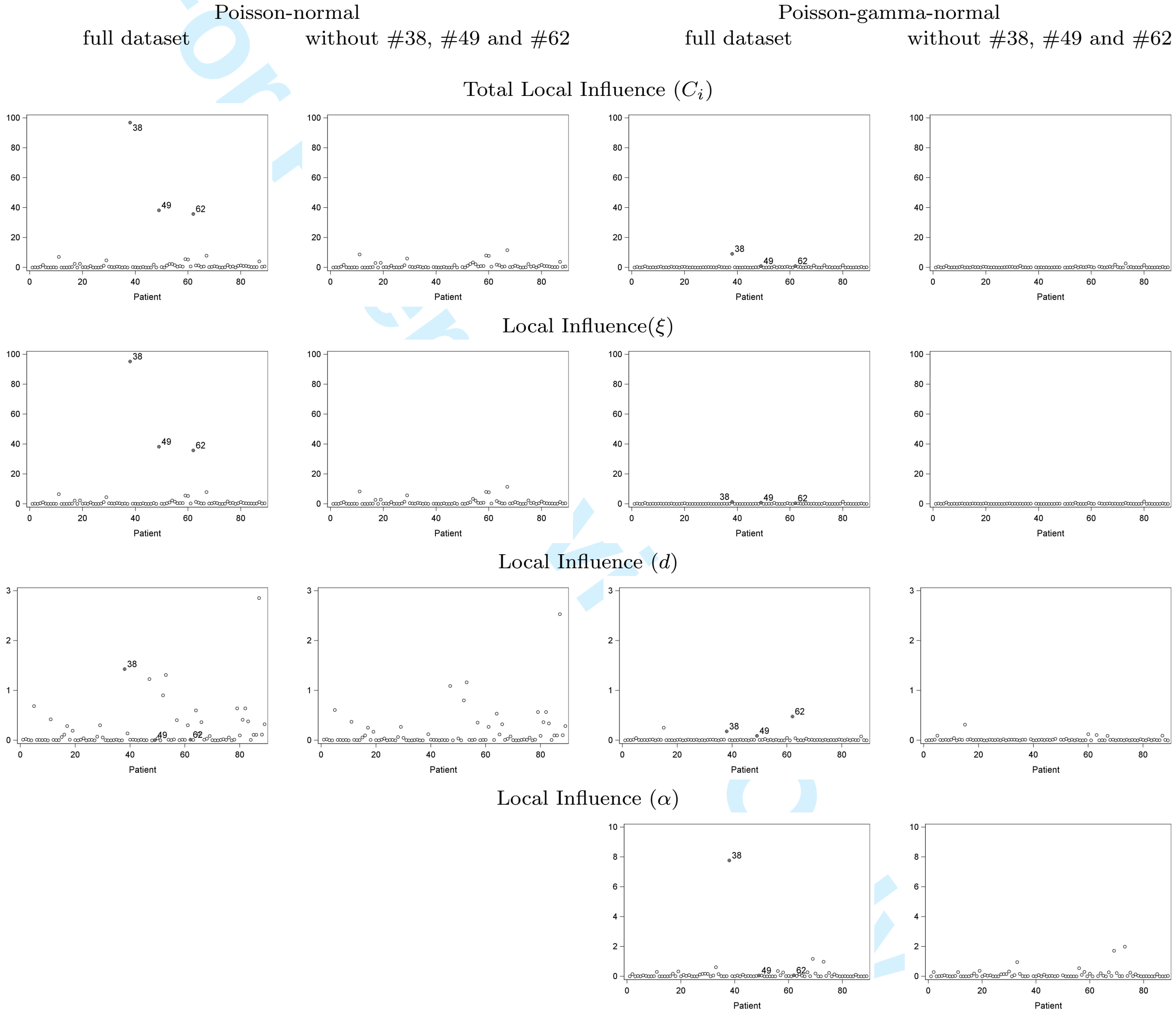

Figure 2. Epilepsy Data. Local influence plots.

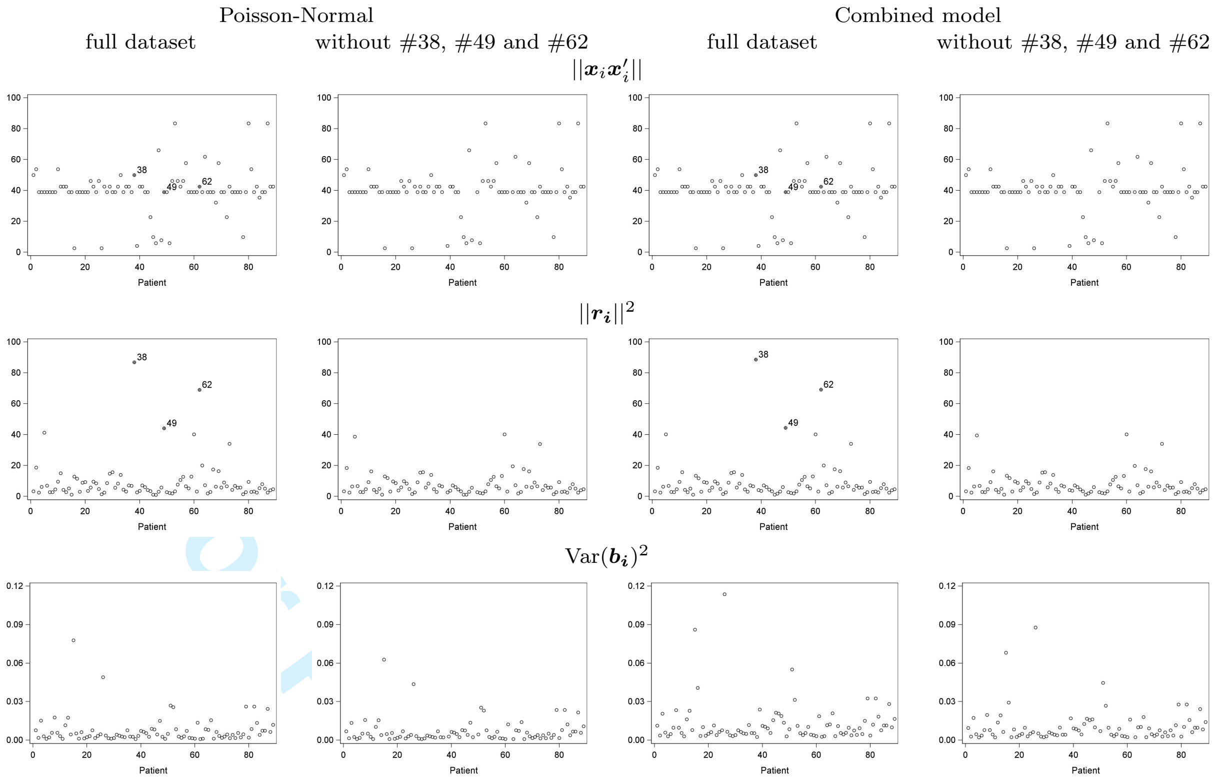

Figure 3. Epilepsy Data. Plots of interpretable components of local influence.

SUPPLEMENTARY MATERIALS

Local Influence Diagnostics for Generalized Linear Mixed Models With Overdispersion

Trias Wahyuni Rakhmawati[a] Geert Molenberghs[a,b] Geert Verbeke[b,a] Christel Faes[a,b]
triaswahyuni.rakhmawati@uhasselt.be

[a] *I-BioStat, Universiteit Hasselt, B-3500 Hasselt, Belgium*
[b] *I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

S.1. Expressions for Standard Local Influence: Subsets

When only a subset $\boldsymbol{\theta}_1$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ is of special interest, the methodology still applies. it follows that (Verbeke and Molenberghs 2000) the corresponding influence:

$$C_h(\boldsymbol{\theta}_1) = C_h \; + \; 2\boldsymbol{h}'\Delta' \begin{pmatrix} 0 & 0 \\ 0 & \ddot{L}_{22}^{-1} \end{pmatrix} \Delta \boldsymbol{h} \leq C_h, \tag{S.1.1}$$

with obvious notation. Should $\ddot{L}_{12} = 0$, then an influence decomposition is possible:

$$C_h = C_h(\boldsymbol{\theta}_1) \; + \; C_h(\boldsymbol{\theta}_2). \tag{S.1.2}$$

For weakly correlated sub-vectors, (S.1.2) holds approximately.

S.2. Local Influence for the Linear Mixed Model

S.2.1 *Standard Approach, Based on the Marginal Likelihood*

The backdrop for our developments is the method as derived for the linear mixed model (Verbeke and Lesaffre 1997b, Verbeke and Molenberghs 2000). We will sketch their developments, and then turn to an alternative derivation based on the likelihood in integral form.

In line with these authors, we consider (4), but with in addition the conditional independence assumption $\Sigma_i = \sigma^2 I_{n_i}$, with I_{n_i} the $n_i \times n_i$ identity matrix.

For C_i as in (21), a convenient form can be derived:

$$C_i = -2\left(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^1_{(i)}\right)' \ddot{L}_{(i)} \ddot{L}^{-1} \ddot{L}_{(i)} \left(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^1_{(i)}\right), \tag{S.2.1}$$

where a subscript (i) indicates that the corresponding quantity is based on the deletion of the ith subject and further the vector $\hat{\boldsymbol{\theta}}^1_{(i)}$ is the one-step approximation to $\hat{\boldsymbol{\theta}}_{(i)}$ obtained from a single Newton-Raphson step in the maximization procedure of $\ell_{(i)}(\boldsymbol{\theta})$, using $\hat{\boldsymbol{\theta}}$ as the starting value. For sufficiently large sample size, it follows that C_i is an approximation to the classical global case-deletion diagnostics. Note that the expression is exact when properly used for local influence purposes.

It is advantageous that C_i admits a closed form (21). Lesaffre and Verbeke (1998) decomposed C_i into five interpretable components. Let $\boldsymbol{\mathcal{R}_i}$, $\mathcal{X}_i$, and $\mathcal{Z}_i$ denote now the "standardized" residuals and covariates for the ith individual, defined by $\boldsymbol{\mathcal{R}_i} = V_i^{-1/2}\boldsymbol{r_i}$, $\mathcal{X}_i = V_i^{-1/2}X_i$, and $\mathcal{Z}_i = V_i^{-1/2}Z_i$, respectively, with $\boldsymbol{r_i} = \boldsymbol{y_i} - X_i\widehat{\boldsymbol{\beta}}$. Further, for a matrix A, let $\|A\| = \sqrt{\text{tr}(A'A)}$ be the Frobenius norm of A (Golub and Van Loan 1989). The interpretable components in C_i are then

$$\|\mathcal{X}_i\mathcal{X}_i{}'\|, \quad \|\boldsymbol{\mathcal{R}_i}\|, \quad \|\mathcal{Z}_i\mathcal{Z}_i{}'\|, \quad \|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|, \quad \|V_i^{-1}\|. \tag{S.2.2}$$

First, $\|\mathcal{X}_i\mathcal{X}_i{}'\|$ measures the "length" of the standardized covariates in the mean structure and $\|\boldsymbol{\mathcal{R}_i}\|$ is an overall measure for how well the observed data for the ith subject are predicted by the mean structure $X_i\boldsymbol{\beta}$. Second, the components $\|\mathcal{Z}_i\mathcal{Z}_i{}'\|$ and $\|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|$ have a similar meaning, but then for the covariance structure. For example, $\|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|$ will be zero only if V_i equals $\boldsymbol{r_i}\boldsymbol{r_i}{}'$. Note that $\boldsymbol{r_i}\boldsymbol{r_i}{}'$ is an estimate for var$(\boldsymbol{y_i})$, which only assumes the mean to be correctly modeled as $X_i\boldsymbol{\beta}$. Therefore, $\|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}{}'\|$ can be interpreted as a residual, capturing how well the covariance structure of the data is modeled by $V_i = Z_iDZ_i' + \sigma^2 I_{n_i}$. Finally, the fifth component $\|V_i^{-1}\|$ will be large if V_i has small eigenvalues, which indicates that the ith subject is assumed to have small variability.

The decomposition of C_i immediately suggests a practical procedure to find an explanation for the influential nature of an individual, i.e., when C_i is large, we examine the diagnostics. Such plots are useful to graphically inspect the individuals in view of their influential nature. Thus, it is sensible to start with an index plot of C_i. Following this, the index plots of (**??**) can be examined. A recurrent practical difficulty with diagnostics is to establish a threshold above which an individual is defined as "remarkable". It follows from (21) that

$$\sum_{i=1}^{N} C_i = -2 \text{ tr}\left(\ddot{L}^{-1}\sum_{i=1}^{N}\boldsymbol{\Delta}_i\boldsymbol{\Delta}_i'\right),$$

which converges to $2s$, for N approaching infinity. As for leverage in linear regression (Neter, Wasserman and Kutner 1990, pp. 395–396), one could classify an individual for which C_i is larger than twice the average value (larger than $4s/N$, for N large) as being influential. However, unlike for the leverage situation, $2s$ is only the approximate sum of the C_i, which will not be accurate if the model is not correctly specified (such that $\ddot{L}^{-1}\sum_i \boldsymbol{\Delta}_i\boldsymbol{\Delta}_i'$ does not converge to I_s) or if N is too small for the asymptotic results to yield good approximations. In such cases, Lesaffre and Verbeke (1998) proposed to replace $2s$ by the actual sum, and we call the ith subject influential if C_i is larger than the cutoff value $2\sum_{i=1}^{\text{N}} C_i/N$.

Given decomposition result (S.1.1), it is interesting to consider sub-vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of fixed effects and variance components, respectively, with corresponding influences $C_i(\boldsymbol{\beta})$ and $C_i(\boldsymbol{\alpha})$, respectively. Given that the fixed effects and variance components are asymptotically independent, it follows that

$$C_i \approx C_i(\boldsymbol{\beta}) \ + \ C_i(\boldsymbol{\alpha}). \tag{S.2.3}$$

Lesaffre and Verbeke (1998) further showed that $C_i(\boldsymbol{\beta})$ can be decomposed using only the first two components $\|\mathcal{X}_i\mathcal{X}_i{}'\|$ and $\|\boldsymbol{\mathcal{R}_i}\|$, while the last three components $\|\mathcal{Z}_i\mathcal{Z}_i{}'\|$,

$\|I - \boldsymbol{\mathcal{R}_i}\boldsymbol{\mathcal{R}_i}'\|$, and $\|V_i^{-1}\|$ feature in the decomposition of $C_i(\boldsymbol{\alpha})$. Asymptotically therefore, influence for the fixed effects and for the variance components can be scrutinized by studying the first two and the last three interpretable components, respectively.

S.2.2 *Integral-based Expression*

As previewed in Section 4, the integral-based approach is used here as an alternative way to alleviate complexities with the explicit marginal likelihood expressions. To prepare for developments of Poisson, probit, logit and Weibull cases, we set out this way for the linear mixed model.

The marginal density corresponding to the linear mixed model is defined by the following expression:

$$\widetilde{f}(\boldsymbol{y}_i) = \int \widetilde{f}(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{b_i})\widetilde{f}(\boldsymbol{b_i}|D)\ d\boldsymbol{b_i}. \tag{S.2.4}$$

The conditional density of the response variable takes the form:

$$\begin{aligned}\widetilde{f}(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{b_i}) &= \left(\frac{1}{2\pi\sigma^2}\right)^{n_i/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b_i})'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b_i})\right\} \\ &= (2\pi s)^{-n_i/2}\exp[f(\boldsymbol{y}_i)],\end{aligned} \tag{S.2.5}$$

where $f(\boldsymbol{y}_i) = -(2s)^{-1}(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)'(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)$; $\widehat{\boldsymbol{y}}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b_i}$, and $s = \sigma^2$. The conditional density of the normal random effect is:

$$f(\boldsymbol{b_i}) = \frac{1}{(2\pi)^{q/2}|D|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}\right) = 2\pi^{-q/2}|D|^{-1/2}\exp\{g(\boldsymbol{b_i})\}, \tag{S.2.6}$$

where $g(\boldsymbol{b_i}) = -\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}$. Thus, the marginal density for the linear mixed model is:

$$\widetilde{f}(\boldsymbol{y}_i) = (2\pi)^{-(n_i+q)/2}s^{-n_i/2}|D|^{-1/2}\int \exp\{f(\boldsymbol{y}_i) + g(\boldsymbol{b_i})\}d\boldsymbol{b_i}. \tag{S.2.7}$$

From (S.2.7) the likelihood derives as:

$$L(\boldsymbol{\beta}, D, s) = \prod_{i=1}^{N} \widetilde{f}(\boldsymbol{y}_i), \tag{S.2.8}$$

and the corresponding log-likelihood is (18). Thus, the log-likelihood contribution of the

ith individual takes the form:

$$\ell_i(\boldsymbol{\beta}, D, s) = \log\left[(2\pi)^{-(n_i+q)/2} s^{-n_i/2} |D|^{-1/2} \int \exp\{f(\boldsymbol{y}_i) + g(\boldsymbol{b_i})\} d\boldsymbol{b_i}\right]$$

$$= -\frac{(n_i+q)}{2}\log(2\pi) - \frac{n_i}{2}\log(s) - \frac{1}{2}\log|D| + \log\int \exp[f(\boldsymbol{y}_i) + g(\boldsymbol{b_i})] d\boldsymbol{b_i}$$

$$\propto -\frac{n_i}{2}\log(s) - \frac{1}{2}\log|D| + \log K_i, \qquad \text{(S.2.9)}$$

where $K_i = \int I_i d\boldsymbol{b_i}$ and $I_i = \exp\{f(\boldsymbol{y}_i) + g(\boldsymbol{b_i})\}$.

To derive the local influence as described in (21), the components of local influence need to be derived. Lesaffre and Verbeke (1998) showed that C_i equals:

$$C_i = 2||\ddot{L}^{-1}||\ ||\boldsymbol{\Delta}_i||^2 \cos(\varphi_i), \qquad \text{(S.2.10)}$$

where φ_i is the angle between $\mathrm{vec}(-\ddot{L}^{-1})$ and $\mathrm{vec}(\boldsymbol{\Delta}_i\boldsymbol{\Delta}_i')$, $\boldsymbol{\Delta}_i$ is the first derivative of $\ell_i(\boldsymbol{\beta}, D, s)$ with respect to the model parameters, and $\ddot{L}^{-1}$ is the $s \times s$ matrix of second derivatives of $\ell(\boldsymbol{\beta}, D, s)$ with respect to the parameters.

The procedure to construct derivatives with respect to the parameters is as follows. First, the derivative with respect to fixed effect $\boldsymbol{\beta}$ is:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D, s)}{\partial \boldsymbol{\beta}} = \frac{1}{K_i}\int I_i \frac{1}{s}\boldsymbol{X_i}'(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i) d\boldsymbol{b_i} = \frac{1}{s}\boldsymbol{X_i}'\frac{\boldsymbol{L_i}}{K_i}, \qquad \text{(S.2.11)}$$

where

$$K_i = \int I_i d\boldsymbol{b_i} = \int \exp[f(\boldsymbol{y}_i) + g(\boldsymbol{b_i})] d\boldsymbol{b_i} = c\widetilde{\phi}(\boldsymbol{y}_i) \qquad \text{(S.2.12)}$$

and

$$\boldsymbol{L_i} = \int I_i(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i) d\boldsymbol{b_i}$$

$$= \int I_i(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b_i}) d\boldsymbol{b_i}$$

$$= \boldsymbol{y}_i \int I_i d\boldsymbol{b_i} - \boldsymbol{X}_i\boldsymbol{\beta}\int I_i d\boldsymbol{b_i} - \boldsymbol{Z}_i \int I_i \boldsymbol{b_i} d\boldsymbol{b_i}. \qquad \text{(S.2.13)}$$

Component $\int I_i \boldsymbol{b_i} d\boldsymbol{b_i}$ of $\boldsymbol{L_i}$ can be rewritten as:

$$\begin{aligned}
\int I_i \boldsymbol{b_i} d\boldsymbol{b_i} &= \int c\widetilde{\phi}(\boldsymbol{y}_i, \boldsymbol{b_i})\boldsymbol{b_i} d\boldsymbol{b_i} \\
&= c\int \widetilde{\phi}(\boldsymbol{y}_i)\widetilde{\phi}(\boldsymbol{b_i}|\boldsymbol{y}_i)\boldsymbol{b_i} d\boldsymbol{b_i} \\
&= c\widetilde{\phi}(\boldsymbol{y}_i)\int \boldsymbol{b_i}\widetilde{\phi}(\boldsymbol{b_i}|\boldsymbol{y}_i) d\boldsymbol{b_i} \\
&= c\widetilde{\phi}(\boldsymbol{y}_i)E(\boldsymbol{b_i}|\boldsymbol{y}_i) \\
&= c\widetilde{\phi}(\boldsymbol{y}_i)D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \\
&= c\widetilde{\phi}(\boldsymbol{y}_i)D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r_i},
\end{aligned} \tag{S.2.14}$$

where $\boldsymbol{r_i} = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$. Expanding the component functions of (S.2.11) leads to:

$$\begin{aligned}
\frac{\partial \ell_i(\boldsymbol{\beta}, D, s)}{\partial \boldsymbol{\beta}} &= \frac{1}{s}\boldsymbol{X_i}'\frac{\boldsymbol{L_i}}{K_i} \\
&= \frac{1}{s}\boldsymbol{X_i}'\left\{\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_iD\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\} \\
&= \frac{1}{s}\boldsymbol{X_i}'\left\{(I_{ni} - \boldsymbol{Z}_iD\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\} \\
&= \frac{1}{s}\boldsymbol{X_i}'\left[\left\{(s + \boldsymbol{Z}_iD\boldsymbol{Z}_i')\boldsymbol{V}_i^{-1} - \boldsymbol{Z}_iD\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\right\}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right] \\
&= \frac{1}{s}\boldsymbol{X_i}'s\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \\
&= \boldsymbol{X_i}'\boldsymbol{V}_i^{-1}\boldsymbol{r_i}.
\end{aligned} \tag{S.2.15}$$

Second, the derivative with respect to $s \equiv \sigma^2$ is as follows:

$$\begin{aligned}
\frac{\partial \ell_i(\boldsymbol{\beta}, D, s)}{\partial s} &= -\frac{n_i}{2s} + \frac{1}{K_i}\int I_i \frac{1}{2s^2}(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)'(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i) d\boldsymbol{b_i} \\
&= -\frac{n_i}{2s} - \frac{1}{sK_i}\int I_i f(\boldsymbol{y}_i) d\boldsymbol{b_i} \\
&= -\frac{1}{s}\left\{\frac{n_i}{2} + \frac{1}{K_i}\int I_i f(\boldsymbol{y}_i) d\boldsymbol{b_i}\right\},
\end{aligned} \tag{S.2.16}$$

where K_i is given in (S.2.12). The component $\int I_i f(\boldsymbol{y}_i) d\boldsymbol{b_i}$ can be rewritten as:

$$\begin{aligned}
\int I_i f(\boldsymbol{y}_i) d\boldsymbol{b_i} &= -\frac{1}{2s}\int I_i(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b_i})'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b_i}) d\boldsymbol{b_i} \\
&= -\frac{1}{2s}\int I_i(\boldsymbol{r_i} - \boldsymbol{Z}_i\boldsymbol{b_i})'(\boldsymbol{r_i} - \boldsymbol{Z}_i\boldsymbol{b_i}) d\boldsymbol{b_i} \\
&= -\frac{1}{2s}\left(\boldsymbol{r_i}'\boldsymbol{r_i}\int I_i d\boldsymbol{b_i} - \boldsymbol{r_i}'\boldsymbol{Z}_i\int \boldsymbol{b_i} I_i d\boldsymbol{b_i}\right) \\
&\qquad\qquad - \frac{1}{2s}\left\{-\left(\int \boldsymbol{b_i} I_i d\boldsymbol{b_i}\right)'\boldsymbol{Z}_i'\boldsymbol{r_i} + \int \boldsymbol{b_i}'\boldsymbol{Z}_i'\boldsymbol{Z}_i\boldsymbol{b_i} I_i d\boldsymbol{b_i}\right\} \\
&= -\frac{1}{2s}c\widetilde{\phi}(\boldsymbol{y}_i)\left\{\boldsymbol{r_i}'\boldsymbol{r_i} - \boldsymbol{r_i}'\boldsymbol{Z}_i E(\boldsymbol{b_i}|\boldsymbol{y_i})\right\} \\
&\quad -\frac{1}{2s}c\widetilde{\phi}(\boldsymbol{y}_i)\left[-(E(\boldsymbol{b_i}|\boldsymbol{y_i}))'\boldsymbol{Z}_i'\boldsymbol{r_i} + E(\boldsymbol{b_i}'\boldsymbol{Z}_i'\boldsymbol{Z}_i\boldsymbol{b_i})\right], \qquad \text{(S.2.17)}
\end{aligned}$$

where $E(\boldsymbol{b_i}|\boldsymbol{y_i}) = D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r_i}$ and

$$\begin{aligned}
E(\boldsymbol{b_i}'\boldsymbol{Z}_i'\boldsymbol{Z}_i\boldsymbol{b_i}) &= \mathrm{tr}\{\boldsymbol{Z}_i\mathrm{Var}(\boldsymbol{b_i}|\boldsymbol{y_i})\boldsymbol{Z}_i'\} + E(\boldsymbol{b_i}|\boldsymbol{y_i})'\boldsymbol{Z}_i'\boldsymbol{Z}_i E(\boldsymbol{b_i}|\boldsymbol{y_i}) \\
&= \mathrm{tr}\left\{\boldsymbol{Z}_i(\boldsymbol{Z}_i's^{-1}\boldsymbol{Z}_i + D^{-1})^{-1}\boldsymbol{Z}_i'\right\} + \boldsymbol{r_i}'\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i D'\boldsymbol{Z}_i'\boldsymbol{Z}_i D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r_i} \\
&= \mathrm{tr}\left[\boldsymbol{Z}_i\left\{D - D\boldsymbol{Z}_i'(s + \boldsymbol{Z}_i D\boldsymbol{Z}_i')^{-1}\boldsymbol{Z}_i D\right\}\boldsymbol{Z}_i'\right] \\
&\qquad + \boldsymbol{r_i}'\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i D'\boldsymbol{Z}_i'\boldsymbol{Z}_i D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r_i} \\
&= \mathrm{tr}\left\{\boldsymbol{Z}_i D\boldsymbol{Z}_i' + (I_{ni} - \boldsymbol{V}_i^{-1}\boldsymbol{Z}_i D\boldsymbol{Z}_i')\right\} \\
&\qquad + \boldsymbol{r_i}'\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i D'\boldsymbol{Z}_i'\boldsymbol{Z}_i D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r_i} \\
&= \mathrm{tr}\left(\boldsymbol{Z}_i D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}s\right) + \boldsymbol{r_i}'\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i D'\boldsymbol{Z}_i'\boldsymbol{Z}_i D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r_i} \\
&= \mathrm{tr}\left\{(\boldsymbol{V}_i - s)\boldsymbol{V}_i^{-1}s\right\} + \boldsymbol{r_i}'\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i D'\boldsymbol{Z}_i'\boldsymbol{Z}_i D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r_i} \\
&= \left\{n_i s - s^2\mathrm{tr}(V_i^{-1})\right\} + \boldsymbol{r_i}'\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i D'\boldsymbol{Z}_i'\boldsymbol{Z}_i D\boldsymbol{Z}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r_i}. \qquad \text{(S.2.18)}
\end{aligned}$$

Thus, (S.2.17) simplifies to:

$$\int I_i f(\boldsymbol{y}_i) d\boldsymbol{b_i} = -\frac{1}{2s} c\widetilde{\phi}(\boldsymbol{y}_i) \left\{ \boldsymbol{r_i}' \left(I_{ni} - \boldsymbol{Z}_i D \boldsymbol{Z}_i' \boldsymbol{V}_i^{-1} - \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i D \boldsymbol{Z}_i' \right) \boldsymbol{r_i} \right\}$$

$$-\frac{1}{2s} c\widetilde{\phi}(\boldsymbol{y}_i) \left\{ \boldsymbol{r_i}' \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i D' \boldsymbol{Z}_i' \boldsymbol{Z}_i D \boldsymbol{Z}_i' \boldsymbol{V}_i^{-1} \boldsymbol{r_i} + n_i s - s^2 \mathrm{tr}(V_i^{-1}) \right\}$$

$$= -\frac{1}{2s} c\widetilde{\phi}(\boldsymbol{y}_i) \left[\boldsymbol{r_i}' \left\{ (I_{ni} - \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i D \boldsymbol{Z}_i')(I_{ni} - \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i D \boldsymbol{Z}_i')' \right\} \boldsymbol{r_i} \right]$$

$$-\frac{1}{2s} c\widetilde{\phi}(\boldsymbol{y}_i) \left\{ n_i s - s^2 \mathrm{tr}(V_i^{-1}) \right\}$$

$$= -\frac{1}{2s} c\widetilde{\phi}(\boldsymbol{y}_i) \left[\boldsymbol{r_i}' M_i M_i' \boldsymbol{r_i} + \left\{ n_i s - s^2 \mathrm{tr}(V_i^{-1}) \right\} \right]$$

$$= -\frac{1}{2s} c\widetilde{\phi}(\boldsymbol{y}_i) \left\{ \boldsymbol{r_i}' V_i^{-1} s s V_i^{-1} \boldsymbol{r_i} + n_i s - s^2 \mathrm{tr}(V_i^{-1}) \right\}, \qquad \text{(S.2.19)}$$

where $M_i = \{V_i^{-1}(s + \boldsymbol{Z}_i D \boldsymbol{Z}_i') - V_i^{-1} \boldsymbol{Z}_i D \boldsymbol{Z}_i'\}$.

Expanding the components of (S.2.16) leads to:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D, s)}{\partial s} = -\frac{1}{s} \left\{ \frac{n_i}{2} + \frac{1}{K_i} \int I_i f(\boldsymbol{y}_i) d\boldsymbol{b_i} \right\}$$

$$= -\frac{n_i}{2s} + \frac{1}{2} \boldsymbol{r_i}' V_i^{-1} V_i^{-1} \boldsymbol{r_i} + \frac{n_i}{2s} - \frac{1}{2} \mathrm{tr}(V_i^{-1})$$

$$= -\frac{1}{2} \left\{ \mathrm{tr}(V_i^{-1}) - \boldsymbol{r_i}' V_i^{-1} V_i^{-1} \boldsymbol{r_i} \right\}. \qquad \text{(S.2.20)}$$

Third, the derivative with respect to D is:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D, s)}{\partial d_{jk}}$$

$$= -\frac{1}{2}(2 - \delta_{jk})(D^{-1})_{jk} + \frac{1}{K_i} \int I_i \frac{\partial g(\boldsymbol{b_i})}{\partial d_{jk}} d\boldsymbol{b_i},$$

$$= -\frac{1}{2}(2 - \delta_{jk})(D^{-1})_{jk} + \frac{1}{2K_i} \int I_i \boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} d\boldsymbol{b_i}, \qquad \text{(S.2.21)}$$

where d_{jk} is the (j, k) element of D. Further, E_{ij} is a matrix of zeros everywhere except a one in entries (j, k) and (k, j). The integral part of the first derivative with respect to D can be written as:

$$\int I_i \boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} d\boldsymbol{b_i} = c\widetilde{\phi}(\boldsymbol{y}_i) \int \boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} \widetilde{\phi}(\boldsymbol{b_i} | \boldsymbol{y}_i) d\boldsymbol{b_i},$$

$$= c\widetilde{\phi}(\boldsymbol{y}_i) E(\boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} | \boldsymbol{y}_i), \qquad \text{(S.2.22)}$$

where

$$E(\boldsymbol{b_i}'D^{-1}E_{jk}D^{-1}\boldsymbol{b_i}|\boldsymbol{y}_i)$$
$$= E\{\mathrm{tr}(\boldsymbol{b_i}'D^{-1}E_{jk}D^{-1}\boldsymbol{b_i})\}$$
$$= \mathrm{tr}[\{D^{-1}E_{jk}D^{-1}E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y}_i)\}]$$
$$= \mathrm{tr}[D^{-1}E_{jk}D^{-1}\left\{\mathrm{Var}(\boldsymbol{b_i}|\boldsymbol{y}_i) + E(\boldsymbol{b_i}|\boldsymbol{y}_i)'E(\boldsymbol{b_i}|\boldsymbol{y}_i)\right\}]$$
$$= \mathrm{tr}\{D^{-1}E_{jk}D^{-1}\mathrm{Var}(\boldsymbol{b_i}|\boldsymbol{y}_i)\} + E(\boldsymbol{b_i}|\boldsymbol{y}_i)'D^{-1}E_{jk}D^{-1}E(\boldsymbol{b_i}|\boldsymbol{y}_i)$$
$$= \mathrm{tr}[D^{-1}E_{jk}D^{-1}\left\{D - D\boldsymbol{Z}_i'(\boldsymbol{s} + \boldsymbol{Z}_iD\boldsymbol{Z}_i')^{-1}\boldsymbol{Z}_iD\right\}] + \boldsymbol{r_i}'V^{-1}\boldsymbol{Z}_iE_{jk}\boldsymbol{Z}_i'V^{-1}\boldsymbol{r_i}$$
$$= \mathrm{tr}\left\{D^{-1}E_{jk}D^{-1}\left(D - D\boldsymbol{Z}_i'V^{-1}\boldsymbol{Z}_iD\right)\right\} + (2-\delta_{jk})\boldsymbol{r_i}'V^{-1}\boldsymbol{Z_i}^{(j)}\boldsymbol{Z_i}^{(k)'}\boldsymbol{Z}_i'V^{-1}\boldsymbol{r_i}$$
$$= \mathrm{tr}(D^{-1}E_{jk}) - \mathrm{tr}(E_{jk}\boldsymbol{Z}_i'V^{-1}\boldsymbol{Z}_iD) + (2-\delta_{jk})\boldsymbol{r_i}'V^{-1}\boldsymbol{Z_i}^{(j)}\boldsymbol{Z_i}^{(k)'}\boldsymbol{Z}_i'V^{-1}\boldsymbol{r_i}$$
$$= (2-\delta_{jk})(D^{-1})_{jk} - (2-\delta_{jk})\boldsymbol{Z_i}^{(j)'}V^{-1}\boldsymbol{Z_i}^{(k)}$$
$$+(2-\delta_{jk})\boldsymbol{r_i}'V^{-1}\boldsymbol{Z_i}^{(j)}\boldsymbol{Z_i}^{(k)'}\boldsymbol{Z}_i'V^{-1}\boldsymbol{r_i}$$
$$= (2-\delta_{jk})\left\{(D^{-1})_{jk} - \boldsymbol{Z_i}^{(j)'}V^{-1}\boldsymbol{Z_i}^{(k)} + \boldsymbol{r_i}'V^{-1}\boldsymbol{Z_i}^{(j)}\boldsymbol{Z_i}^{(k)'}\boldsymbol{Z}_i'V^{-1}\boldsymbol{r_i}\right\}. \quad \text{(S.2.23)}$$

Expanding the components of (S.2.21) leads to:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D, s)}{\partial d_{jk}}$$
$$= -\frac{1}{2}(2-\delta_{jk})\left(\boldsymbol{Z_i}^{(j)'}V^{-1}\boldsymbol{Z_i}^{(k)} - \boldsymbol{r_i}'V^{-1}\boldsymbol{Z_i}^{(j)}\boldsymbol{Z_i}^{(k)'}\boldsymbol{Z}_i'V^{-1}\boldsymbol{r_i}\right). \quad \text{(S.2.24)}$$

This integral-based result, based on (S.2.15), (S.2.20), and (S.2.24) is identical to the standard one of Lesaffre and Verbeke (1998). Hence also, the same interpretable components as in (S.2.2) ensue.

S.3. Local Influence for the Poisson-normal Model

S.3.1 *Explicit Marginal Expression*

The log-likelihood contribution for the ith subject based on closed form solution (7) is defined as:

$$\ell_i(\boldsymbol{\beta}, D) = \log\left(\frac{1}{\prod_{j=1}^{n_i} y_{ij}!}\right) + \log\sum_{\boldsymbol{t}} \frac{(-1)^{\sum_{j=1}^{n_i} t_j}}{\prod_{j=1}^{n_i} t_j!} \cdot \exp\left\{\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{x}_{ij}'\boldsymbol{\xi}\right\}$$
$$\times \exp\left[\frac{1}{2}\left\{\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}_{ij}'\right\} D \left\{\sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}_{ij}\right\}\right]$$
$$\propto \log\left\{\sum_{\boldsymbol{t}} \frac{(-1)^{\sum_{j=1}^{n_i} t_j}}{\prod_{j=1}^{n_i} t_j!} \cdot \exp\left(\boldsymbol{V}_{\boldsymbol{t}_i}{}'\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{W}_{\boldsymbol{t}_i}{}'D\boldsymbol{W}_{\boldsymbol{t}_i}\right)\right\}$$
$$\propto \log\left(\sum_{\boldsymbol{t}} K_{t_i}\right), \tag{S.3.1}$$

where

$$\boldsymbol{V}_{\boldsymbol{t}_i} = \sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{x}_{ij},$$
$$\boldsymbol{W}_{\boldsymbol{t}_i} = \sum_{j=1}^{n_i}(y_{ij}+t_j)\boldsymbol{z}_{ij},$$
$$K_{t_i} = \frac{(-1)^{\sum_{j=1}^{n_i} t_j}}{\prod_{j=1}^{n_i} t_j!} \cdot \exp\left(\boldsymbol{V}_{\boldsymbol{t}_i}{}'\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{W}_{\boldsymbol{t}_i}{}'D\boldsymbol{W}_{\boldsymbol{t}_i}\right).$$

The first derivative with respect to the fixed effects is:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} = \frac{1}{\sum_{\boldsymbol{t}} K_{t_i}} \sum_{\boldsymbol{t}} \frac{(-1)^{\sum_{j=1}^{n_i} t_j}}{\prod_{j=1}^{n_i} t_j!} \cdot \exp\left(\boldsymbol{V}_{\boldsymbol{t}_i}{}'\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{W}_{\boldsymbol{t}_i}{}'D\boldsymbol{W}_{\boldsymbol{t}_i}\right)\boldsymbol{V}_{\boldsymbol{t}_i},$$
$$= \frac{\sum_{\boldsymbol{t}} K_{t_i}\boldsymbol{V}_{\boldsymbol{t}_i}}{\sum_{\boldsymbol{t}} K_{t_i}}. \tag{S.3.2}$$

Further,the first derivative with respect to the D matrix components is:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = \frac{1}{\sum_{\boldsymbol{t}} K_{t_i}} \sum_{\boldsymbol{t}} K_{t_i}\left(\frac{1}{2}\boldsymbol{W}_{\boldsymbol{t}_i}{}'DE_{jk}\boldsymbol{W}_{\boldsymbol{t}_i}\right), \tag{S.3.3}$$

where E_{ij} is as in Section 5.1.2. The infinite series of $\boldsymbol{t}$ in all first derivatives is as defined in (S.3.2) and (S.3.3); it may lead to computational difficulties. Hence, the integral-based expression here is more promising, as will be demonstrated next.

S.3.2 *Integral-based Expression*

To circumvent the occurrence of infinite series as in the previous section, the likelihood based on integral approach was used to derive the interpretable components of local influence. From the marginal likelihood in integral form (3), the joint marginal density for the Poisson-normal model could be written:

$$
\begin{aligned}
\mathrm{P}(\boldsymbol{Y_i}=\boldsymbol{y_i}) &= \int \prod_{j=1}^{n_i} \frac{1}{y_{ij}!} \exp\{(\gamma_{ij}+c_{ij})y_{ij}\} \exp[-\exp\{(\gamma_{ij}+c_{ij})\}] \\
&\qquad \times \frac{\exp(-\frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i})}{(2\pi)^{q/2}|D|^{1/2}} d\boldsymbol{b_i} \\
&= \left\{ \prod_{j=1}^{n_i} \frac{1}{y_{ij}!} \frac{\exp(\gamma_{ij}y_{ij})}{(2\pi)^{q/2}|D|^{1/2}} \right\} \\
&\qquad \times \int \prod_{j=1}^{n_i} \exp\left\{ c_{ij}y_{ij} - \exp(\gamma_{ij}+c_{ij}) - \frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i} \right\} d\boldsymbol{b_i} \\
&= \frac{1}{(2\pi)^{q/2}|D|^{1/2}} \left(\prod_{j=1}^{n_i} \frac{1}{y_{ij}!} \right) \exp\left\{ \sum_{j=1}^{n_i} (\gamma_{ij}y_{ij}) \right\} I_i, \qquad \text{(S.3.4)}
\end{aligned}
$$

where $\gamma_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ and $c_{ij} = \boldsymbol{z}'_{ij}\boldsymbol{b_i}$. Further, I_i is as follows:

$$
\begin{aligned}
I_i &= \int \prod_{j=1}^{n_i} \exp\left\{ c_{ij}y_{ij} - \exp(\gamma_{ij}+c_{ij}) - \frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i} \right\} d\boldsymbol{b_i} \\
&= \int \exp\left[\sum_{j=1}^{n_i} \{c_{ij}y_{ij} - \exp(\gamma_{ij}+c_{ij})\} - \frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i} \right] d\boldsymbol{b_i} \\
&= \int \exp\left\{ \widetilde{f}(y_i) + \widetilde{f}(b_i) \right\} d\boldsymbol{b_i} = \int A_i d\boldsymbol{b_i} = c\widetilde{\eta}(\boldsymbol{y_i}),
\end{aligned}
$$

where:

$$
\widetilde{f}(\boldsymbol{y_i}) = \sum_{j=1}^{n_i} \{c_{ij}y_{ij} - \exp(\gamma_{ij}+c_{ij})\},
$$

$$
\widetilde{f}(\boldsymbol{b_i}) = -\frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i},
$$

$$
A_i = \exp\left\{ \widetilde{f}(\boldsymbol{y_i}) + \widetilde{f}(\boldsymbol{b_i}) \right\}.
$$

From (S.3.4), the log-likelihood contribution for subject i derives as:

$$\ell_i(\boldsymbol{\beta}, D) = -\frac{q}{2}\log(2\pi) - \frac{1}{2}\log|D| + \log\left(\prod_{j=1}^{n_i}\frac{1}{y_{ij}!}\right) + \sum_{j=1}^{n_i}(\gamma_{ij}y_{ij}) + \log I_i$$

$$\propto -\frac{1}{2}\log|D| + \sum_{j=1}^{n_i}(\gamma_{ij}y_{ij}) + \log I_i. \quad \text{(S.3.5)}$$

First, we construct the first derivative with respect to the fixed-effect parameter:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n_i}\boldsymbol{x}_{ij}y_{ij} + \frac{1}{I_i}\frac{\partial I_i}{\partial \boldsymbol{\beta}}. \quad \text{(S.3.6)}$$

The first derivative of I_i can be rewritten as:

$$\frac{\partial I_i}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} \int \prod_{j=1}^{n_i} \exp\left\{c_{ij}y_{ij} - \exp(\gamma_{ij} + c_{ij}) - \frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i}\right\} d\boldsymbol{b_i}$$

$$= \int \frac{\partial}{\partial \boldsymbol{\beta}} \prod_{j=1}^{n_i} \exp(c_{ij}y_{ij}) \exp\{-\exp(\gamma_{ij} + c_{ij})\} \exp\left(-\frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i}\right) d\boldsymbol{b_i}$$

$$= \int \left[\frac{\partial}{\partial \boldsymbol{\beta}} \exp\left\{-\sum_{j=1}^{n_i} \exp(\gamma_{ij} + c_{ij})\right\}\right] \exp\left\{\sum_{j=1}^{n_i}(c_{ij}y_{ij}) - \frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i}\right\} d\boldsymbol{b_i}$$

$$= \int \left[\exp\left\{-\sum_{j=1}^{n_i} \exp(\gamma_{ij} + c_{ij})\right\} \cdot \frac{\partial}{\partial \boldsymbol{\beta}} \left\{-\sum_{j=1}^{n_i} \exp(\gamma_{ij} + c_{ij})\right\}\right]$$
$$\times \exp\left\{\sum_{j=1}^{n_i}(c_{ij}y_{ij}) - \frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i}\right\} d\boldsymbol{b_i}$$

$$= \int \left[\exp\left\{-\sum_{j=1}^{n_i} \exp(\gamma_{ij} + c_{ij})\right\} \cdot \left\{-\sum_{j=1}^{n_i} \exp(\gamma_{ij} + c_{ij})\boldsymbol{x}_{ij}\right\}\right]$$
$$\times \exp\left\{\sum_{j=1}^{n_i}(c_{ij}y_{ij}) - \frac{1}{2}\boldsymbol{b_i}' D^{-1}\boldsymbol{b_i}\right\} d\boldsymbol{b_i}$$

$$= -\int A_i \sum_{j=1}^{n_i} \exp(\gamma_{ij} + c_{ij})\boldsymbol{x}_{ij} d\boldsymbol{b_i}$$

$$= -c\widetilde{\eta}(\boldsymbol{y_i}) \int \widetilde{\eta}(\boldsymbol{b_i}|\boldsymbol{y_i}) \sum_{j=1}^{n_i} \exp(\gamma_{ij} + c_{ij})\boldsymbol{x}_{ij} d\boldsymbol{b_i}$$

$$= -c\widetilde{\eta}(\boldsymbol{y_i}) \sum_{j=1}^{n_i} \left\{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\boldsymbol{x}_{ij} \int \exp(\boldsymbol{z}'_{ij}\boldsymbol{b_i})\widetilde{\eta}(\boldsymbol{b_i}|\boldsymbol{y_i}) d\boldsymbol{b_i}\right\}$$

$$= -c\widetilde{\eta}(\boldsymbol{y_i}) \sum_{j=1}^{n_i} \left[\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\boldsymbol{x}_{ij} E\{\exp(\boldsymbol{z}'_{ij}\boldsymbol{b_i}|\boldsymbol{y_i})\}\right]$$

$$= -c\widetilde{\eta}(\boldsymbol{y_i}) \sum_{j=1}^{n_i} \left\{\exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}'_{ij} D \boldsymbol{z}_{ij}\right) \boldsymbol{x}_{ij}\right\}. \qquad \text{(S.3.7)}$$

Expanding the component function of (S.3.6) leads to:

$$\begin{aligned}\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} &= \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} y_{ij} + \frac{1}{I_i}\frac{\partial I_i}{\partial \boldsymbol{\beta}} \\ &= \sum_{j=1}^{n_i} \left\{ \boldsymbol{x}_{ij} y_{ij} - \exp\left(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij} \right) \boldsymbol{x}_{ij} \right\} \\ &= \sum_{j=1}^{n_i} \left\{ y_{ij} - \exp\left(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij} \right) \right\} \boldsymbol{x}_{ij} \\ &= \sum_{j=1}^{n_i} \left\{ y_{ij} - E(y_{ij}|\boldsymbol{b_i}) \right\} \boldsymbol{x}_{ij} \\ &= \sum_{j=1}^{n_i} r_{ij} \boldsymbol{x}_{ij}. \end{aligned} \tag{S.3.8}$$

Second, the first derivative with respect to the D components is:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = -\frac{1}{2}(2 - \delta_{jk})(D^{-1})_{jk} + \frac{1}{I_i}\frac{\partial I_i}{\partial d_{jk}}. \tag{S.3.9}$$

The first derivative of I_i with respect to D can be derived as:

$$\begin{aligned}\frac{\partial I_i}{\partial d_{jk}} &= \int \exp \widetilde{f}(y_i) \frac{\partial}{\partial d_{jk}} \exp \widetilde{f}(b_i) d\boldsymbol{b_i} \\ &= \int \exp \widetilde{f}(y_i) \exp \widetilde{f}(b_i) \left\{ -\frac{1}{2}\boldsymbol{b_i}'(-D^{-1}E_{jk}D^{-1})\boldsymbol{b_i} \right\} d\boldsymbol{b_i} \\ &= \frac{1}{2}\int A_i \boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} d\boldsymbol{b_i} \\ &= \frac{1}{2} c\widetilde{\eta}(\boldsymbol{y_i}) \int \widetilde{\eta}(\boldsymbol{b_i}|\boldsymbol{y_i}) \boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} d\boldsymbol{b_i} \\ &= \frac{1}{2} c\widetilde{\eta}(\boldsymbol{y_i}) E(\boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i}) \\ &= \frac{1}{2} c\widetilde{\eta}(\boldsymbol{y_i}) E\left\{ \mathrm{tr}(\boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i}) \right\} \\ &= \frac{1}{2} c\widetilde{\eta}(\boldsymbol{y_i}) \mathrm{tr}\left\{ D^{-1} E_{jk} D^{-1} E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i}) \right\}. \end{aligned} \tag{S.3.10}$$

The quantity $E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i})$ is derived by means of the closed-form solution for the Poisson-normal model, as follows:

$$
\begin{aligned}
E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i}) &= \int \boldsymbol{b_i}'\boldsymbol{b_i}\widetilde{\eta}(\boldsymbol{b_i}|\boldsymbol{y_i})d\boldsymbol{b_i} \\
&= \int \boldsymbol{b_i}'\boldsymbol{b_i}\frac{\widetilde{\eta}(\boldsymbol{y_i}|\boldsymbol{b_i})\widetilde{\eta}(\boldsymbol{b_i})}{\widetilde{\eta}(\boldsymbol{y_i})}d\boldsymbol{b_i} \\
&= \int \boldsymbol{b_i}'\boldsymbol{b_i}\frac{\sum_{\boldsymbol{t}} F_{ij}\exp(\boldsymbol{\omega}'_{ij}\boldsymbol{b_i} - \frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i})}{|D|^{1/2}(2\pi)^{q/2}} \cdot \frac{1}{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})}d\boldsymbol{b_i} \\
&= \frac{1}{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})} \cdot \frac{1}{|D|^{1/2}(2\pi)^{q/2}} \\
&\quad \times \sum_{\boldsymbol{t}} F_{ij}\int \boldsymbol{b_i}'\boldsymbol{b_i}\exp\left(\boldsymbol{\omega}'_{ij}\boldsymbol{b_i} - \frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}\right)d\boldsymbol{b_i}, \qquad \text{(S.3.11)}
\end{aligned}
$$

where

$$
F_{ij} = \sum_{\boldsymbol{t}} \frac{(-1)^{\sum_{j=1}^{n_i} t_j}}{\prod_{j=1}^{n_i} t_j!} \cdot \exp\left\{\sum_{j=1}^{n_i}(y_{ij} + t_j)\boldsymbol{x}'_{ij}\boldsymbol{\xi}\right\} \qquad \text{(S.3.12)}
$$

and $\boldsymbol{\omega}_{ij} = \sum_{j=1}^{n_i}(y_{ij} + t_j)\boldsymbol{z}_{ij}$. We now reorganize the components of the exponential expression under the integral form in (S.3.11) as follows:

$$
-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i} + \boldsymbol{\omega}'_{ij}\boldsymbol{b_i} = -\frac{1}{2}(\boldsymbol{b_i} - \boldsymbol{k})'D^{-1}(\boldsymbol{b_i} - \boldsymbol{k}) + \boldsymbol{\ell}, \qquad \text{(S.3.13)}
$$

where

$$
\boldsymbol{k} = D\boldsymbol{\omega}_{ij}, \qquad \boldsymbol{\ell} = \frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij}, \qquad \widetilde{\boldsymbol{b_i}} = \boldsymbol{b_i} - \boldsymbol{k}.
$$

Combining (S.3.11) and (S.3.13) produces:

$$\begin{aligned}
E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i}) &= \frac{\sum_{\boldsymbol{t}} F_{ij}}{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})} \cdot \frac{1}{|D|^{1/2}(2\pi)^{q/2}} \\
&\qquad \times \int (\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k})'(\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k}) \exp\left(-\frac{1}{2}\widetilde{\boldsymbol{b}}_{\boldsymbol{i}}{}' D^{-1}\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{\ell}\right) d\boldsymbol{b_i} \\
&= \frac{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})}{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})} \cdot \frac{1}{|D|^{1/2}(2\pi)^{q/2}} \\
&\qquad \times \int (\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k})'(\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k}) \exp\left(-\frac{1}{2}\widetilde{\boldsymbol{b}}_{\boldsymbol{i}}{}' D^{-1}\widetilde{\boldsymbol{b}}_{\boldsymbol{i}}\right) d\boldsymbol{b_i} \\
&= \frac{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})}{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})} \cdot E\left\{(\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k})'(\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k})\right\} \\
&= \frac{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})}{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})} \cdot E\left\{\widetilde{\boldsymbol{b}}_{\boldsymbol{i}}{}'\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + 2\widetilde{\boldsymbol{b}}_{\boldsymbol{i}}\boldsymbol{k} + \boldsymbol{k}'\boldsymbol{k}\right\} \\
&= \frac{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})}{\sum_{\boldsymbol{t}} F_{ij}\exp(\frac{1}{2}\boldsymbol{\omega}'_{ij}D\boldsymbol{\omega}_{ij})} \cdot E\left(\boldsymbol{b_i}'\boldsymbol{b_i}\right) \\
&= \mathrm{Var}(\boldsymbol{b_i}). \qquad \text{(S.3.14)}
\end{aligned}$$

Plugging (S.3.14) into (S.3.10) yields:

$$\begin{aligned}
\frac{\partial I_i}{\partial d_{jk}} &= \frac{1}{2}c\widetilde{\eta}(\boldsymbol{y_i})\mathrm{tr}\left\{D^{-1}E_{jk}D^{-1}E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i})\right\} \\
&= \frac{1}{2}c\widetilde{\eta}(\boldsymbol{y_i})\mathrm{tr}\left\{D^{-1}E_{jk}D^{-1}\mathrm{Var}(\boldsymbol{b_i})\right\} \\
&= \frac{1}{2}c\widetilde{\eta}(\boldsymbol{y_i})\mathrm{tr}\left(D^{-1}E_{jk}D^{-1}\right)\mathrm{Var}(\boldsymbol{b_i}) \\
&= \frac{1}{2}c\widetilde{\eta}(\boldsymbol{y_i})(2-\delta_{jk})(D^{-1}D^{-1})_{jk}\mathrm{Var}(\boldsymbol{b_i}). \qquad \text{(S.3.15)}
\end{aligned}$$

Thus, (S.3.9) leads to:

$$\begin{aligned}
\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} &= -\frac{1}{2}(2-\delta_{jk})(D^{-1})_{jk} + \frac{1}{I_i}\frac{\partial I_i}{\partial d_{jk}} \\
&= -\frac{1}{2}(2-\delta_{jk})\left\{(D^{-1})_{jk} - (D^{-1}D^{-1})_{jk}\mathrm{Var}(\boldsymbol{b_i})\right\}, \qquad \text{(S.3.16)}
\end{aligned}$$

where δ_{jk} is one if j is equal to k, and zero otherwise. We can rewrite the local influence expression as in (S.2.10), where $\boldsymbol{\Delta}_i$, the first-order partial derivative of the contribution of the ith subject to the log-likelihood, is given by:

$$\boldsymbol{\Delta}_i = \begin{bmatrix} \sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij} \\ -\frac{1}{2}(D^{-1})_{11} + \frac{1}{2}(D^{-1}D^{-1})_{11}\mathrm{Var}(\boldsymbol{b_i}) \\ -(D^{-1})_{12} + (D^{-1}D^{-1})_{12}\mathrm{Var}(\boldsymbol{b_i}) \\ -\frac{1}{2}(D^{-1})_{22} + \frac{1}{2}(D^{-1}D^{-1})_{22}\mathrm{Var}(\boldsymbol{b_i}) \\ \cdot \\ \cdot \\ \cdot \\ -(D^{-1})_{q-1,q} + (D^{-1}D^{-1})_{q-1,q}\mathrm{Var}(\boldsymbol{b_i}) \\ -\frac{1}{2}(D^{-1})_{qq} + \frac{1}{2}(D^{-1}D^{-1})_{qq}\mathrm{Var}(\boldsymbol{b_i}) \end{bmatrix}.$$

Rewriting $||\boldsymbol{\Delta}_i||^2$ as the sum of the squares of the contributions for the ith individual yields:

$$\begin{aligned} ||\boldsymbol{\Delta}_i||^2 &= \left(\sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}\right)\left(\sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}\right)' \\ &+ \sum_{k=1}^{q}\left\{-\frac{1}{2}(D^{-1})_{kk} + \frac{1}{2}(D^{-1}D^{-1})_{kk}\mathrm{Var}(\boldsymbol{b_i})\right\}^2 \\ &+ \sum_{k<l}\left\{-(D^{-1})_{kl} + (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}^2 \\ &= \left(\sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}\right)\left(\sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}\right)' \\ &+ \sum_{k,l}\left\{-\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}^2. \end{aligned} \quad \text{(S.3.17)}$$

Let $C_i = C_{1i} + C_{2i}$ with:

$$C_{1i} = 2||\ddot{L}^{-1}||\ ||\boldsymbol{r_i}\boldsymbol{x}_i||^2\cos(\varphi_i), \quad \text{(S.3.18)}$$

$$C_{2i} = \frac{1}{2}||\ddot{L}^{-1}||\ ||(D^{-1})_{kl} - (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})||^2\cos(\varphi_i), \quad \text{(S.3.19)}$$

where $\boldsymbol{r_i}\boldsymbol{x}_i = \sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}$. Note that C_{1i} and C_{2i} are the contributions of subject i to local influence C_i from $\boldsymbol{\beta}$ and D, respectively. Reconstructing the component C_{1i} leads to:

$$||\boldsymbol{r_i}@_i||^2 = \mathrm{tr}(\boldsymbol{r_i}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{r_i}') = \mathrm{vec}(\boldsymbol{x}_i\boldsymbol{x}_i')\mathrm{vec}(\boldsymbol{r_i}\boldsymbol{r_i}') = \cos(\alpha_i)||\boldsymbol{x}_i\boldsymbol{x}_i'||\ ||\boldsymbol{r_i}||^2, \quad \text{(S.3.20)}$$

where $\cos(\alpha_i)$ is the angle between $\mathrm{vec}(\boldsymbol{x}_i\boldsymbol{x}_i')$ and $\mathrm{vec}(\boldsymbol{r_i}\boldsymbol{r_i}')$. Further, the component C_{2i}

is:

$$\begin{aligned}&||(D^{-1})_{kl} - (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})||^2\\&= \mathrm{tr}\left[\left\{(D^{-1})_{kl} - (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}^2\right]\\&= \mathrm{tr}\left\{(D^{-1})^2_{kl}\right\} - \mathrm{tr}\left\{2(D^{-1})_{kl}(D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}\\&\quad + \mathrm{tr}\left\{(D^{-1}D^{-1})^2_{kl}\mathrm{Var}(\boldsymbol{b_i})^2\right\}. \qquad \text{(S.3.21)}\end{aligned}$$

Thus, it follows from (S.3.20) and (S.3.21) that:

$$C_{1i} = 2||\ddot{L}^{-1}||\ ||\boldsymbol{x}_i\boldsymbol{x}_i'||\ ||\boldsymbol{r_i}||^2\cos(\alpha_i)\cos(\varphi_i), \qquad \text{(S.3.22)}$$

$$\begin{aligned}C_{2i} &= \frac{1}{2}||\ddot{L}^{-1}||\cos(\varphi_i)\\&\times\left[\mathrm{tr}\left\{(D^{-1})^2_{kl}\right\} - \mathrm{tr}\left\{2(D^{-1})_{kl}(D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}\right.\\&\quad\left.+\mathrm{tr}\left\{(D^{-1}D^{-1})^2_{kl}\mathrm{Var}(\boldsymbol{b_i})^2\right\}\right]. \qquad \text{(S.3.23)}\end{aligned}$$

Hence, the interpretable components of C_i in the case of the Poisson-normal model can be described using the 'length of the fixed effect' $||\boldsymbol{x}_i\boldsymbol{x}_i'||$, the 'squared length of the residual' $||\boldsymbol{r_i}||^2$, and the 'squared of random effect variability' $\mathrm{Var}(\boldsymbol{b_i})^2$.

S.4. Local Influence for the Probit-normal Model

Given that the numerical route was generically described in Section 5.1.3, we will focus on the explicit calculations only, using the integral method. The general binary case of the logit-based model can be described as:

$$\mathrm{P}(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{b_i}) = \prod_{j=1}^{n_i}\lambda_{ij}^{y_{ij}}(1-\lambda_{ij})^{(1-y_{ij})}, \qquad \text{(S.4.1)}$$

where κ_{ij} is (10). The marginal joint density function between probit response and normal random effect takes the following form:

$$\begin{aligned}&f(\boldsymbol{y}_i = 1)\\&= \frac{1}{(2\pi)^{q/2}|D|^{1/2}}\int\left\{\prod_{j=1}^{n_i}\Phi_{n_i}(\boldsymbol{X}_i'\boldsymbol{\xi} + \boldsymbol{Z}_i'\boldsymbol{b_i})\right\}\exp\left(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}\right)d\boldsymbol{b_i}. \qquad \text{(S.4.2)}\end{aligned}$$

Above expression, as explained in Molenberghs (2010) leads to (11). The log-likehood contribution for the ith subject is

$$\ell_i(\boldsymbol{\beta}, D) \propto \frac{1}{2}\log|L| - \frac{1}{2}\log I_i,$$

where $\boldsymbol{s} = \boldsymbol{X}_i\boldsymbol{\beta}$, $u = -\frac{1}{2}\boldsymbol{s}'L\boldsymbol{s}$ and

$$\begin{aligned} I_i &= \int_{-\infty}^{\boldsymbol{X}_i\boldsymbol{\beta}} \exp\left(-\frac{1}{2}\boldsymbol{s}'L\boldsymbol{s}\right) d\boldsymbol{s} \\ &= \int_{-\infty}^{\boldsymbol{X}_i\boldsymbol{\beta}} \exp(u)(\boldsymbol{s}^{-1})'L^{-1}du \\ &= \{\exp(\boldsymbol{X}_i\boldsymbol{\beta})\} \cdot \{(\boldsymbol{X}_i\boldsymbol{\beta})^{-1}\}'L^{-1}. \end{aligned}$$

The first derivative with respect to the fixed effects is

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} = \frac{1}{I_i}\frac{\partial I_i}{\partial \boldsymbol{\beta}} = [I - (\boldsymbol{X}_i\boldsymbol{\beta})^{-1}]\boldsymbol{X}_i.$$

Further, the derivative with respect to D is

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = \frac{1}{2L}\frac{\partial L}{\partial d_{jk}} + \frac{1}{I_i}\frac{\partial I_i}{\partial d_{jk}}, \tag{S.4.3}$$

where

$$\begin{aligned} \frac{\partial L}{\partial d_{jk}} &= \frac{\partial}{\partial d_{jk}}\left\{I_{n_i} - Z_i\left(D^{-1} + Z_i'Z_i\right)^{-1} Z_i'\right\} \\ &= I_{n_i} - Z_iM_iM_i'D^{-1}E_{jk}D^{-1}Z_i' \\ &= I_{n_i} - Z_iM_iM_i'(D^{-1}D^{-1})_{jk}Z_i', \end{aligned} \tag{S.4.4}$$

and

$$\begin{aligned} \frac{\partial I_i}{\partial d_{jk}} &= \frac{\partial}{\partial d_{jk}}\left[\{\exp(\boldsymbol{X}_i\boldsymbol{\beta})\} \cdot \left\{(\boldsymbol{X}_i\boldsymbol{\beta})^{-1}\right\}' L^{-1}\right] \\ &= \{\exp(\boldsymbol{X}_i\boldsymbol{\beta})\} \cdot \left\{(\boldsymbol{X}_i\boldsymbol{\beta})^{-1}\right\}' L^{-1}L^{-1}\frac{\partial L}{\partial d_{jk}} \\ &= \{\exp(\boldsymbol{X}_i\boldsymbol{\beta})\} \cdot \left\{(\boldsymbol{X}_i\boldsymbol{\beta})^{-1}\right\}' L^{-1}L^{-1} \\ &\quad \times \left\{I_{n_i} - Z_iM_iM_i'(D^{-1}D^{-1})_{jk}Z_i'\right\}, \end{aligned} \tag{S.4.5}$$

where $M_i = \left(D^{-1} + Z_i'Z_i\right)^{-1}$. Thus, (S.4.3) takes the following form:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = \frac{3}{2}L^{-1}\left\{I_{n_i} - Z_iM_iM_i'(D^{-1}D^{-1})_{jk}Z_i'\right\}. \tag{S.4.6}$$

The first-order derivative for ith subject yields:

$$\begin{aligned} ||\boldsymbol{\Delta}_i||^2 &= \{I - (\boldsymbol{X}_i\boldsymbol{\beta})^{-1}\}^2\boldsymbol{X}_i\boldsymbol{X}_i' \\ &\quad + \sum_{k,l}\frac{9}{4}L^{-2}\left\{I_{n_i} - Z_iM_iM_i'(D^{-1}D^{-1})_{jk}Z_i'\right\}^2 \end{aligned} \tag{S.4.7}$$

Hence, reformulating above expression leads to the components of local influence for the probit-normal model: $||\boldsymbol{X_i}||^2$ and $||\boldsymbol{Z_i}\boldsymbol{Z_i}'||^2$.

S.5. Local Influence for the Logit-normal Model

In this section, we will derive local influence using the integral-based approach for the logit-normal model. The general binary case of logit model follows from:

$$\mathrm{P}(\boldsymbol{y}_i|\boldsymbol{\beta},\boldsymbol{b_i}) = \prod_{j=1}^{n_i} \lambda_{ij}^{y_{ij}}(1-\lambda_{ij})^{(1-y_{ij})}, \tag{S.5.1}$$

where

$$\kappa_{ij} = \frac{\exp(\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b_i})}{1+\exp(\boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b_i})}.$$

The marginal joint density function between logit response and normal random effect takes the following form:

$$f(\boldsymbol{y}_i = 1) = \frac{1}{(2\pi)^{q/2}|D|^{1/2}} \int \prod_{j=1}^{n_i} \lambda_{ij} \exp(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i})d\boldsymbol{b_i}. \tag{S.5.2}$$

The log-likelihood contribution for the ith subject is

$$\ell_i(\boldsymbol{\beta}, D) = \log\left\{\frac{1}{(2\pi)^{q/2}|D|^{1/2}}\right\} + \log\left\{\int \prod_{j=1}^{n_i} \lambda_{ij} \exp\left(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}\right) d\boldsymbol{b_i}\right\}$$

$$\propto -\frac{1}{2}\log|D| + \log I_i, \tag{S.5.3}$$

where $I_i = \int J_i d\boldsymbol{b_i} = c\tilde{\tau}(\boldsymbol{y}_i)$ and

$$J_i = \prod_{j=1}^{n_i} \lambda_{ij} \exp\left(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}\right). \tag{S.5.4}$$

The first derivative with respect to fixed effects:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} = \frac{1}{I_i}\int \frac{\partial}{\partial \boldsymbol{\beta}}\left(\prod_{j=1}^{n_i}\lambda_{ij}\right)\cdot \exp\left(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}\right) d\boldsymbol{b_i}$$

$$= \frac{1}{I_i}\int \left(\prod_{j=1}^{n_i}\lambda_{ij}\right)\left(\sum_{j=1}^{n_i}\frac{1}{\lambda_{ij}}\frac{\partial \lambda_{ij}}{\partial \boldsymbol{\beta}}\right) \exp\left(-\frac{1}{2}\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}\right) d\boldsymbol{b_i}, \tag{S.5.5}$$

where

$$\frac{\partial \lambda_{ij}}{\partial \boldsymbol{\beta}} = \frac{\exp(\mu_{ij})\boldsymbol{x}_{ij}\{1+\exp(\mu_{ij})\} - \exp(\mu_{ij})\exp(\mu_{ij})\boldsymbol{x}_{ij}}{\{1+\exp(\mu_{ij})\}^2} = \frac{\lambda_{ij}\boldsymbol{x}_{ij}}{1+\exp(\mu_{ij})}.$$

Thus, equation (S.5.5) leads to:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} = \frac{1}{I_i} \int J_i \left(\sum_{j=1}^{n_i} \frac{\boldsymbol{x}_{ij}}{1+\exp(\mu_{ij})} \right) d\boldsymbol{b_i}$$

$$= \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \int \frac{1}{1+\exp(\mu_{ij})} \tilde{\tau}(\boldsymbol{b_i}|\boldsymbol{y}_i) d\boldsymbol{b_i}, \qquad \text{(S.5.6)}$$

where $\mu_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}$.

Further, the first derivative with respect to D is:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = -\frac{1}{2}(2-\delta_{jk})(D^{-1})_{jk} + \frac{1}{2I_i} \int J_i \boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} d\boldsymbol{b_i}. \qquad \text{(S.5.7)}$$

Expanding the integral expression of (S.5.7) leads the same result as (S.3.14) and (S.6.12). Thus, the first derivative with respect to D takes the following form:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = -\frac{1}{2}(2-\delta_{jk})(D^{-1})_{jk} + \frac{1}{2}\text{tr}\left\{D^{-1}E_{jk}D^{-1}\text{Var}(\boldsymbol{b_i})\right\}$$

$$= -\frac{1}{2}(2-\delta_{jk})\left\{(D^{-1})_{jk} - (D^{-1}D^{-1})_{jk}\text{Var}(\boldsymbol{b_i})\right\}. \qquad \text{(S.5.8)}$$

The vector $\boldsymbol{\Delta}_i$ of first-order partial derivative of the contribution of the ith subject to the log-likelihood is now given by:

$$\boldsymbol{\Delta}_i = \begin{bmatrix} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \int \frac{1}{1+\exp(\mu_{ij})} \tilde{\tau}(\boldsymbol{b_i}|\boldsymbol{y}_i) d\boldsymbol{b_i} \\ -\frac{1}{2}(D^{-1})_{11} + \frac{1}{2}(D^{-1}D^{-1})_{11}\text{Var}(\boldsymbol{b_i}) \\ -(D^{-1})_{12} + (D^{-1}D^{-1})_{12}\text{Var}(\boldsymbol{b_i}) \\ -\frac{1}{2}(D^{-1})_{22} + \frac{1}{2}(D^{-1}D^{-1})_{22}\text{Var}(\boldsymbol{b_i}) \\ \cdot \\ \cdot \\ \cdot \\ -(D^{-1})_{q-1,q} + (D^{-1}D^{-1})_{q-1,q}\text{Var}(\boldsymbol{b_i}) \\ -\frac{1}{2}(D^{-1})_{qq} + \frac{1}{2}(D^{-1}D^{-1})_{qq}\text{Var}(\boldsymbol{b_i}) \end{bmatrix}.$$

Rewriting $||\boldsymbol{\Delta}_i||^2$ as the sum of squares of the contributions for the ith individual yields:

$$||\boldsymbol{\Delta}_i||^2 = \left\{ \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \int \frac{1}{1+\exp(\mu_{ij})} \tilde{\tau}(\boldsymbol{b_i}|\boldsymbol{y}_i) d\boldsymbol{b_i} \right\}^2$$

$$+ \sum_{k=1}^{q} \left\{ -\frac{1}{2}(D^{-1})_{kk} + \frac{1}{2}(D^{-1}D^{-1})_{kk} \mathrm{Var}(\boldsymbol{b_i}) \right\}^2$$

$$+ \sum_{k<l} \left\{ -(D^{-1})_{kl} + (D^{-1}D^{-1})_{kl} \mathrm{Var}(\boldsymbol{b_i}) \right\}^2$$

$$\propto \left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \right) \left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \right)'$$

$$+ \sum_{k,l} \left\{ -\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl} \mathrm{Var}(\boldsymbol{b_i}) \right\}^2 . \tag{S.5.9}$$

Reconstructing the components of fixed- and random-effects as in the Poisson and Weibull cases, leads to:

$$C_{1i} = 2||\ddot{L}^{-1}|| \; ||\boldsymbol{x}_i||^2 \cos(\varphi_i), \tag{S.5.10}$$

$$C_{2i} = \frac{1}{2}||\ddot{L}^{-1}|| \cos(\varphi_i)$$

$$\times \left[\mathrm{tr}\left\{ (D^{-1})^2_{kl} \right\} - \mathrm{tr}\left\{ 2(D^{-1})_{kl}(D^{-1}D^{-1})_{kl} \mathrm{Var}(\boldsymbol{b_i}) \right\} + \mathrm{tr}\left\{ (D^{-1}D^{-1})^2_{kl} \mathrm{Var}(\boldsymbol{b_i})^2 \right\} \right] . \tag{S.5.11}$$

Hence, the interpretable components of C_i for the logit-normal model can be described using the length of fixed effect $||\boldsymbol{x}_i||^2$ and the squared of random effect variability $\mathrm{Var}(\boldsymbol{b_i})^2$, in analogy with the Poisson-normal and Weibull-normal models.

S.6. Local Influence for the Weibull-normal Model

The general Weibull model for repeated measurement data as described in (14) can be re-expressed as:

$$f(\boldsymbol{y}_i|\boldsymbol{\theta}_i, \boldsymbol{b_i}) = \prod_{j=1}^{n_i} \lambda \rho y_{ij}^{\rho-1} \exp(\boldsymbol{\mu_{ij}}) \exp\{-\lambda y_{ij}^{\rho} \exp(\boldsymbol{\mu_{ij}})\}$$

$$= \lambda\rho \left(\prod_{j=1}^{n_i} y_{ij}^{\rho-1} \right) \exp\left[\sum_{j=1}^{n_i} \left\{ \boldsymbol{\mu_{ij}} - \lambda y_{ij}^{\rho} \exp(\boldsymbol{\mu_{ij}}) \right\} \right]$$

$$= \lambda\rho \left(\prod_{j=1}^{n_i} y_{ij}^{\rho-1} \right) \exp\left\{ \tilde{f}(\boldsymbol{y_i}) \right\}, \tag{S.6.1}$$

where $\boldsymbol{\mu_{ij}} = \boldsymbol{x}'_{ij}\boldsymbol{\xi} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}$ and $\widetilde{f}(\boldsymbol{y_i}) = \sum_{j=1}^{n_i}\left\{\boldsymbol{\mu_{ij}} - \lambda y_{ij}^{\rho}\exp(\boldsymbol{\mu_{ij}})\right\}$. Thus, the marginal density of the Weibull-model takes the following form:

$$\begin{aligned} f(\boldsymbol{y}_i) &= \int f(\boldsymbol{y}_i|\boldsymbol{\beta},\boldsymbol{b_i})f(\boldsymbol{b_i}|D)\ d\boldsymbol{b_i} \\ &= \frac{\lambda\rho\left(\prod_{j=1}^{n_i} y_{ij}^{\rho-1}\right)}{(2\pi)^{q/2}|D|^{1/2}}\int \exp[\widetilde{f}(\boldsymbol{y_i}) + \widetilde{g}(\boldsymbol{b_i})]d\boldsymbol{b_i}, \end{aligned} \tag{S.6.2}$$

where $\widetilde{g}(\boldsymbol{b_i}) = -\boldsymbol{b_i}'D^{-1}\boldsymbol{b_i}/2$. The log-likelihood contribution for the ith subject can be written as:

$$\begin{aligned} \ell_i(\boldsymbol{\beta}, D) &= \log\left[\frac{\lambda\rho\left(\prod_{j=1}^{n_i} y_{ij}^{\rho-1}\right)}{(2\pi)^{q/2}|D|^{1/2}}\int \exp\{\widetilde{f}(\boldsymbol{y_i}) + \widetilde{g}(\boldsymbol{b_i})\}d\boldsymbol{b_i}\right] \\ &\propto -\frac{1}{2}\log|D| + \log K_i, \end{aligned} \tag{S.6.3}$$

where $K_i = \int I_i d\boldsymbol{b_i} = c\widetilde{\phi}(\boldsymbol{y_i})$ and $I_i = \exp\{\widetilde{f}(\boldsymbol{y_i}) + \widetilde{g}(\boldsymbol{b_i})\}$.

The first derivative of the log-likelihood with respect to the fixed effects takes the following form:

$$\begin{aligned} \frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} &= \frac{1}{K_i}\int I_i\left[\sum_{j=1}^{n_i}\{\boldsymbol{x}_{ij} - \lambda y_{ij}^{\rho}\exp(\boldsymbol{\mu_{ij}})\boldsymbol{x}_{ij}]\right\}d\boldsymbol{b_i} \\ &= \sum_{j=1}^{n_i}\boldsymbol{x}_{ij} - \frac{1}{K_i}\sum_{j=1}^{n_i}\left\{\lambda y_{ij}^{\rho}\boldsymbol{x}_{ij}\int I_i\exp(\boldsymbol{\mu_{ij}})d\boldsymbol{b_i}\right\}. \end{aligned} \tag{S.6.4}$$

The component relative to the integral part in (S.6.4) can be rewritten as:

$$\begin{aligned} \int I_i\exp(\boldsymbol{\mu_{ij}})d\boldsymbol{b_i} &= c\widetilde{\phi}(\boldsymbol{y_i})\int \exp(\boldsymbol{\mu_{ij}})\widetilde{\phi}(\boldsymbol{b_i}|\boldsymbol{y_i})d\boldsymbol{b_i} \\ &= c\widetilde{\phi}(\boldsymbol{y_i})\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\int \exp(\boldsymbol{z}'_i\boldsymbol{b_i})\widetilde{\phi}(\boldsymbol{b_i}|\boldsymbol{y_i})d\boldsymbol{b_i} \\ &= c\widetilde{\phi}(\boldsymbol{y_i})\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\exp\left(\frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}\right) \\ &= c\widetilde{\phi}(\boldsymbol{y_i})\exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}'_{ij}D\boldsymbol{z}_{ij}\right). \end{aligned} \tag{S.6.5}$$

Expanding the component functions in (S.6.4) leads to:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n_i}\boldsymbol{x}_{ij} - \lambda\sum_{j=1}^{n_i} y_{ij}^{\rho}\boldsymbol{x}_{ij}\exp(\boldsymbol{\mu_{ij}}). \tag{S.6.6}$$

Further, the first derivative with respect to the D-components is:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = -\frac{1}{2}(2-\delta_{jk})(D^{-1})_{jk} + \frac{1}{2K_i}\int I_i \boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} d\boldsymbol{b_i}. \tag{S.6.7}$$

Solving the integral expression leads to:

$$\begin{aligned}\int I_i \boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} d\boldsymbol{b_i} &= c\widetilde{\phi}(\boldsymbol{y_i}) E(\boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i} | \boldsymbol{y_i}) \\ &= c\widetilde{\phi}(\boldsymbol{y_i}) E\{\mathrm{tr}(\boldsymbol{b_i}' D^{-1} E_{jk} D^{-1} \boldsymbol{b_i})\} \\ &= c\widetilde{\phi}(\boldsymbol{y_i}) \mathrm{tr}\{D^{-1} E_{jk} D^{-1} E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i})\},\end{aligned} \tag{S.6.8}$$

where E_{jk} is as in Section 5.1.2. Expectation $E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i})$ is derived using the closed form for the Weibull-normal model:

$$\begin{aligned}&E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i}) \\ &= \int \boldsymbol{b_i}'\boldsymbol{b_i} \widetilde{\phi}(\boldsymbol{b_i}|\boldsymbol{y_i}) d\boldsymbol{b_i} \\ &= \int \boldsymbol{b_i}'\boldsymbol{b_i} \frac{\widetilde{\phi}(\boldsymbol{y_i}|\boldsymbol{b_i})\widetilde{\eta}(\boldsymbol{b_i})}{\widetilde{\phi}(\boldsymbol{y_i})} d\boldsymbol{b_i} \\ &= \frac{1}{\widetilde{\phi}(\boldsymbol{y_i})} \int \boldsymbol{b_i}'\boldsymbol{b_i} \sum_{\boldsymbol{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \\ &\quad \times \frac{1}{|D|^{1/2}(2\pi)^{q/2}} \exp\left\{(m_j+1)\boldsymbol{\mu}_{ij} - \frac{1}{2}\boldsymbol{b_i}' D^{-1} \boldsymbol{b_i}\right\} d\boldsymbol{b_i} \\ &= \frac{1}{\widetilde{\phi}(\boldsymbol{y_i})} \sum_{\boldsymbol{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \\ &\quad \times \frac{1}{|D|^{1/2}(2\pi)^{q/2}} \int \boldsymbol{b_i}'\boldsymbol{b_i} \exp\left\{(m_j+1)\boldsymbol{\mu}_{ij} - \frac{1}{2}\boldsymbol{b_i}' D^{-1} \boldsymbol{b_i}\right\} d\boldsymbol{b_i},\end{aligned} \tag{S.6.9}$$

where $\widetilde{\phi}(\boldsymbol{y_i})$ equals (15). Reorganizing the components of the exponential expression in the integrand of (S.6.9) leadas to:

$$-\frac{1}{2}\boldsymbol{b_i}' D^{-1} \boldsymbol{b_i} + (m_j+1)\boldsymbol{\mu}_{ij} = -\frac{1}{2}(\boldsymbol{b_i} - \boldsymbol{k})' D^{-1} (\boldsymbol{b_i} - \boldsymbol{k}) + \boldsymbol{\ell}, \tag{S.6.10}$$

with

$$\boldsymbol{k} = (m_j+1) D \boldsymbol{z}_{ij}, \qquad \boldsymbol{\ell} = (m_j+1)[\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \frac{1}{2}(m_j+1)\boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}], \qquad \widetilde{\boldsymbol{b_i}} = \boldsymbol{b_i} - \boldsymbol{k}.$$

Rewriting $\widetilde{\phi}(\boldsymbol{y_i})$ leads to:

$$\widetilde{\phi}(\boldsymbol{y_i}) = \sum_{\boldsymbol{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \exp(\boldsymbol{\ell}). \tag{S.6.11}$$

Combining (S.6.9) and (S.6.10) produces:

$$\begin{aligned} E(\boldsymbol{b_i}'\boldsymbol{b_i}|\boldsymbol{y_i}) &= \frac{1}{\widetilde{\phi}(\boldsymbol{y_i})} \sum_{\boldsymbol{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \cdot \frac{\exp(\boldsymbol{\ell})}{|D|^{1/2}(2\pi)^{q/2}} \\ &\quad \times \int (\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k})'(\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k}) \exp\left(-\frac{1}{2}\widetilde{\boldsymbol{b}}_{\boldsymbol{i}}' D^{-1} \widetilde{\boldsymbol{b}}_{\boldsymbol{i}}\right) d\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} \\ &= \frac{1}{\widetilde{\phi}(\boldsymbol{y_i})} \sum_{\boldsymbol{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \exp(\boldsymbol{\ell}) E\left\{(\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k})'(\widetilde{\boldsymbol{b}}_{\boldsymbol{i}} + \boldsymbol{k})\right\} \\ &= \frac{1}{\widetilde{\phi}(\boldsymbol{y_i})} \sum_{\boldsymbol{m}} \prod_{j=1}^{n_i} \frac{(-1)^{m_j}}{m_j!} \lambda^{m_j+1} \rho y_{ij}^{(m_j+1)\rho-1} \exp(\boldsymbol{\ell}) E\left(\boldsymbol{b_i}'\boldsymbol{b_i}\right) \\ &= \mathrm{Var}(\boldsymbol{b_i}). \end{aligned} \tag{S.6.12}$$

Plugging (S.6.12) into (S.6.8) and (S.6.7) yields:

$$\begin{aligned} \frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} &= -\frac{1}{2}(2-\delta_{jk})(D^{-1})_{jk} + \frac{1}{2}\mathrm{tr}\left\{D^{-1}E_{jk}D^{-1}\mathrm{Var}(\boldsymbol{b_i})\right\} \\ &= -\frac{1}{2}(2-\delta_{jk})\left\{(D^{-1})_{jk} - (D^{-1}D^{-1})_{jk}\mathrm{Var}(\boldsymbol{b_i})\right\}, \end{aligned} \tag{S.6.13}$$

where δ_{jk} is as before.

The vector $\boldsymbol{\Delta}_i$ of first-order partial derivative of the contribution of the ith subject to the log-likelihood is now given by:

$$\boldsymbol{\Delta}_i = \begin{bmatrix} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} - \lambda \sum_{j=1}^{n_i} y_{ij}^{\rho} \boldsymbol{x}_{ij} \exp(\boldsymbol{\mu_{ij}}) \\ -\frac{1}{2}(D^{-1})_{11} + \frac{1}{2}(D^{-1}D^{-1})_{11}\mathrm{Var}(\boldsymbol{b_i}) \\ -(D^{-1})_{12} + (D^{-1}D^{-1})_{12}\mathrm{Var}(\boldsymbol{b_i}) \\ -\frac{1}{2}(D^{-1})_{22} + \frac{1}{2}(D^{-1}D^{-1})_{22}\mathrm{Var}(\boldsymbol{b_i}) \\ \cdot \\ \cdot \\ \cdot \\ -(D^{-1})_{q-1,q} + (D^{-1}D^{-1})_{q-1,q}\mathrm{Var}(\boldsymbol{b_i}) \\ -\frac{1}{2}(D^{-1})_{qq} + \frac{1}{2}(D^{-1}D^{-1})_{qq}\mathrm{Var}(\boldsymbol{b_i}) \end{bmatrix}.$$

Rewriting $||\boldsymbol{\Delta}_i||^2$ as the sum of squares of the contributions for the ith individual yields:

$$||\boldsymbol{\Delta}_i||^2 = \left\{\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} - \boldsymbol{Q}_i\right\}^2 + \sum_{k=1}^{q}\left\{-\frac{1}{2}(D^{-1})_{kk} + \frac{1}{2}(D^{-1}D^{-1})_{kk}\mathrm{Var}(\boldsymbol{b_i})\right\}^2$$
$$+\sum_{k<l}\{-(D^{-1})_{kl} + (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\}^2$$
$$= \left(\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}\right)\left(\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}\right)' - 2\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}\boldsymbol{Q}_i' + \boldsymbol{Q}_i\boldsymbol{Q}_i'$$
$$+\sum_{k,l}\left\{-\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}^2, \quad \text{(S.6.14)}$$

where $\boldsymbol{Q}_i = \lambda\sum_{j=1}^{n_i} y_{ij}^{\rho}\boldsymbol{x}_{ij}\exp(\boldsymbol{\mu_{ij}})$. Write $C_i = C_{1i} + C_{2i}$, with:

$$C_{1i} = 2||\ddot{L}^{-1}||\ \left\{||\boldsymbol{x}_i||^2 - 2\boldsymbol{x}_i\boldsymbol{Q}_i + ||\boldsymbol{Q}_i||^2\right\}\cos(\varphi_i), \quad \text{(S.6.15)}$$

$$C_{2i} = \frac{1}{2}||\ddot{L}^{-1}||\ ||(D^{-1})_{kl} - (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})||^2\cos(\varphi_i), \quad \text{(S.6.16)}$$

where $\boldsymbol{x}_i = \sum_{j=1}^{n_i}\boldsymbol{x}_{ij}$. Note that C_{1i} and C_{2i} are the contributions of the ith subject to local influence contributions C_i from $\boldsymbol{\beta}$ and D, respectively. Rewriting the component of C_{2i} leads to:

$$||(D^{-1})_{kl} - (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})||^2$$
$$= \mathrm{tr}\left[\left\{(D^{-1})_{kl} - (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}^2\right]$$
$$= \mathrm{tr}\left\{(D^{-1})_{kl}^2\right\} - \mathrm{tr}\left\{2(D^{-1})_{kl}(D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}$$
$$+\ \mathrm{tr}\left\{(D^{-1}D^{-1})_{kl}^2\mathrm{Var}(\boldsymbol{b_i})^2\right\}. \quad \text{(S.6.17)}$$

It then follows that:

$$C_{1i} = 2||\ddot{L}^{-1}||\ \left(||\boldsymbol{x}_i||^2 - 2\boldsymbol{x}_i\boldsymbol{Q}_i + ||\boldsymbol{Q}_i||^2\right)\cos(\varphi_i), \quad \text{(S.6.18)}$$

$$C_{2i} = \frac{1}{2}||\ddot{L}^{-1}||\cos(\varphi_i)$$
$$\times\left[\mathrm{tr}\left\{(D^{-1})_{kl}^2\right\} - \mathrm{tr}\left\{2(D^{-1})_{kl}(D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b_i})\right\}\right.$$
$$\left.+\mathrm{tr}\left\{(D^{-1}D^{-1})_{kl}^2\mathrm{Var}(\boldsymbol{b_i})^2\right\}\right]. \quad \text{(S.6.19)}$$

Hence, the interpretable components of C_i for the Weibull normal model can be described using the length of fixed effect ($||\boldsymbol{x}_i||^2$) and the squared of random effect variability ($\mathrm{Var}(\boldsymbol{b_i})^2$), in analogy with the Poisson-normal model.

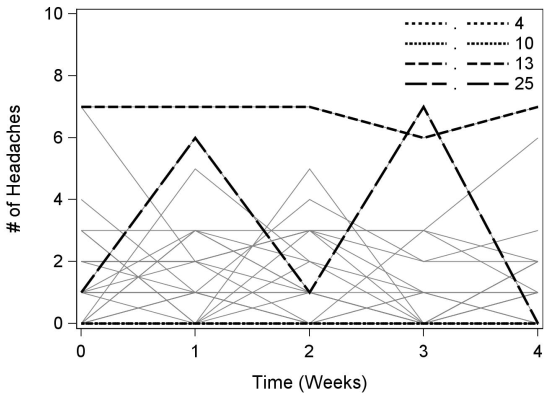

Figure S.4. Headache Data. Individual profiles.

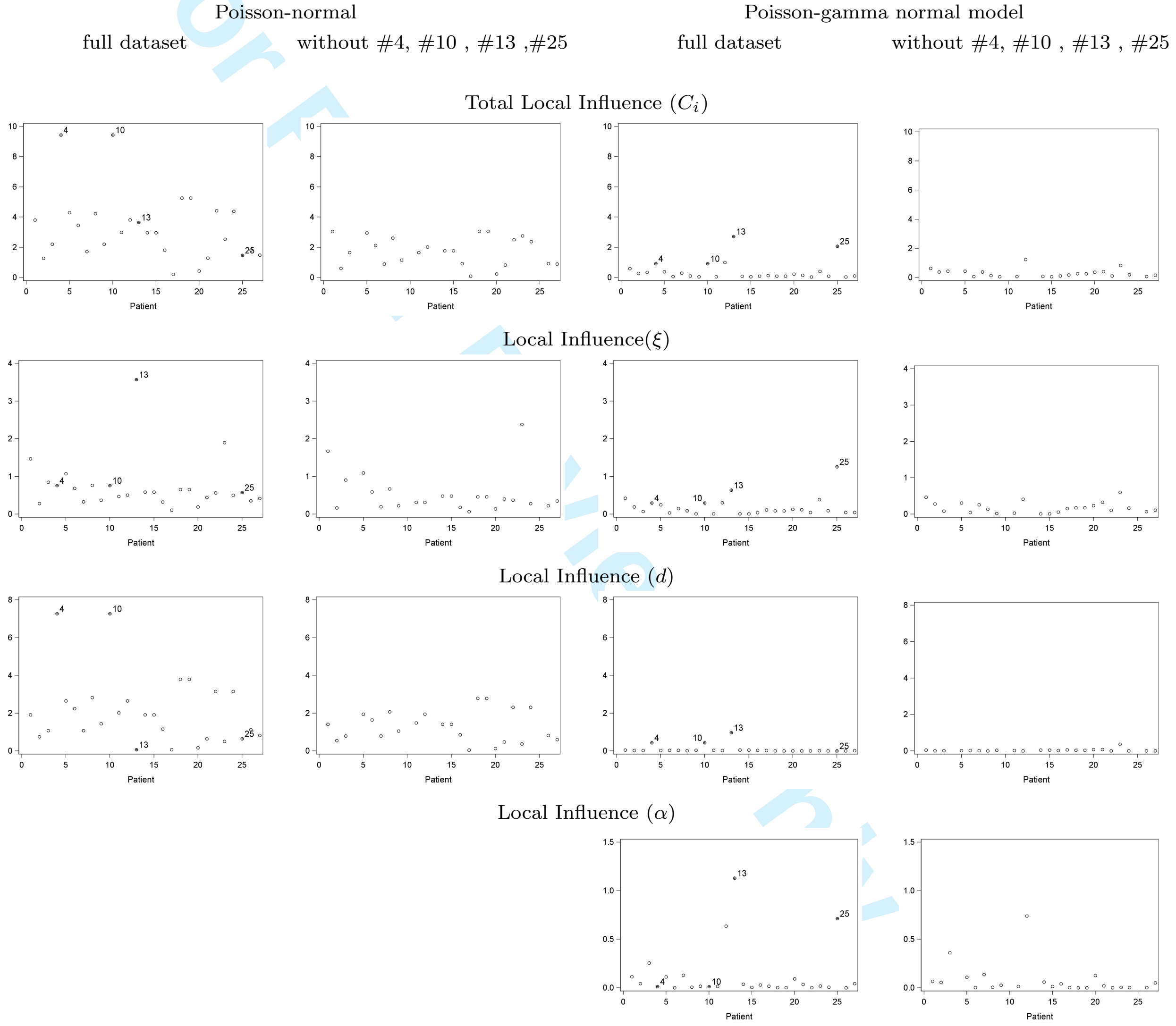

Figure S.5. Headache Data. Local influence plots.

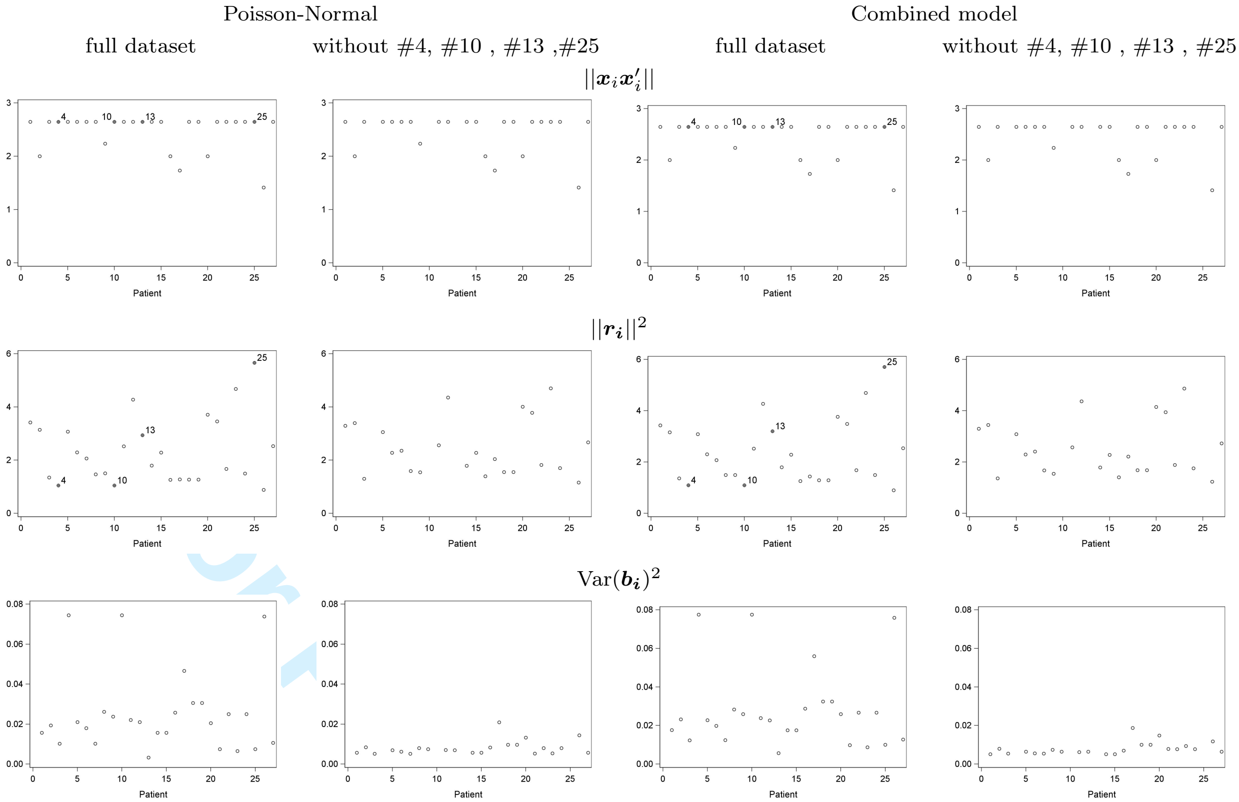

Figure S.6. Headache Data. Plot of interpretable components of local influence.

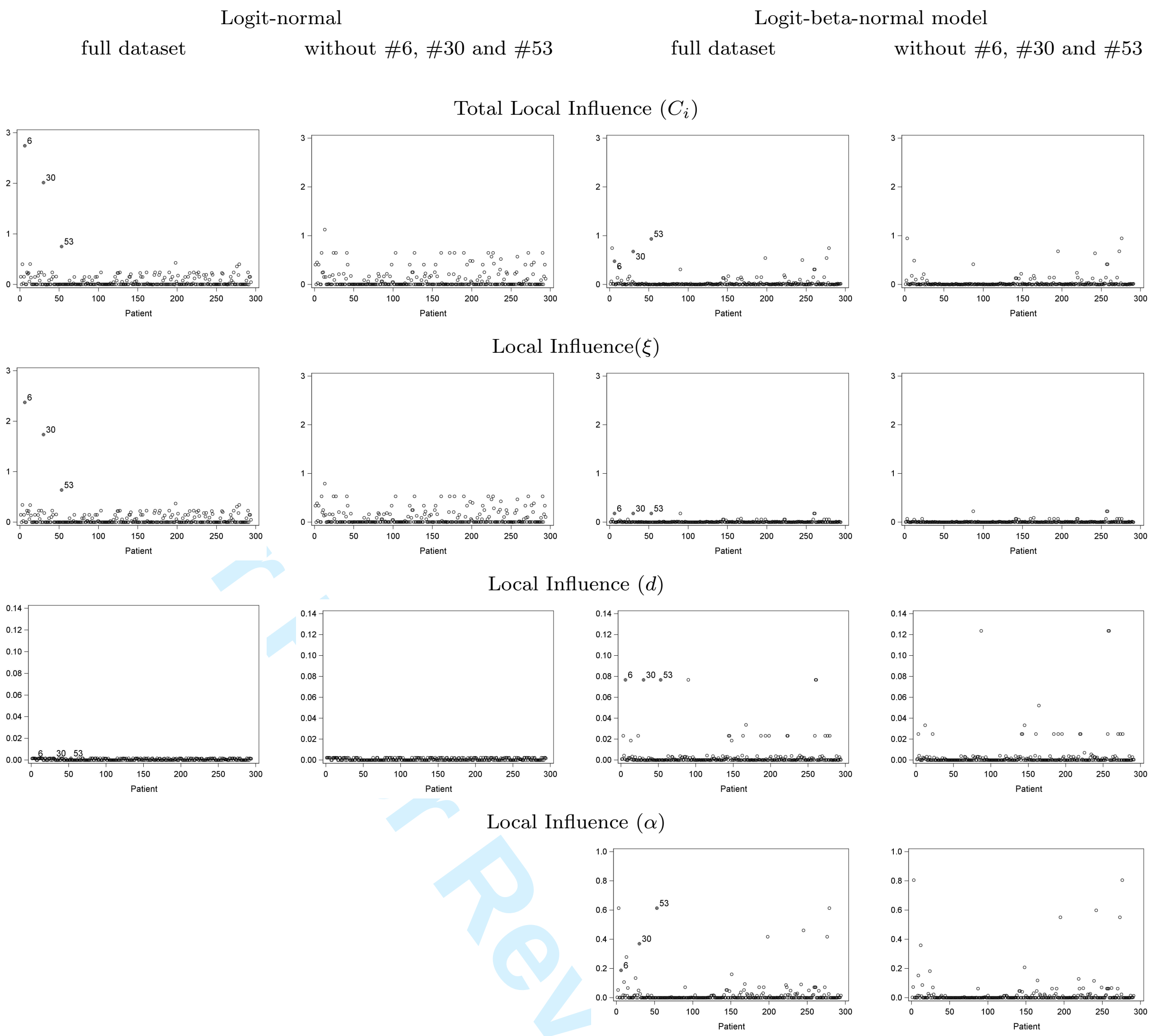

Figure S.7. Onychomycosis Data. Local influence plots.

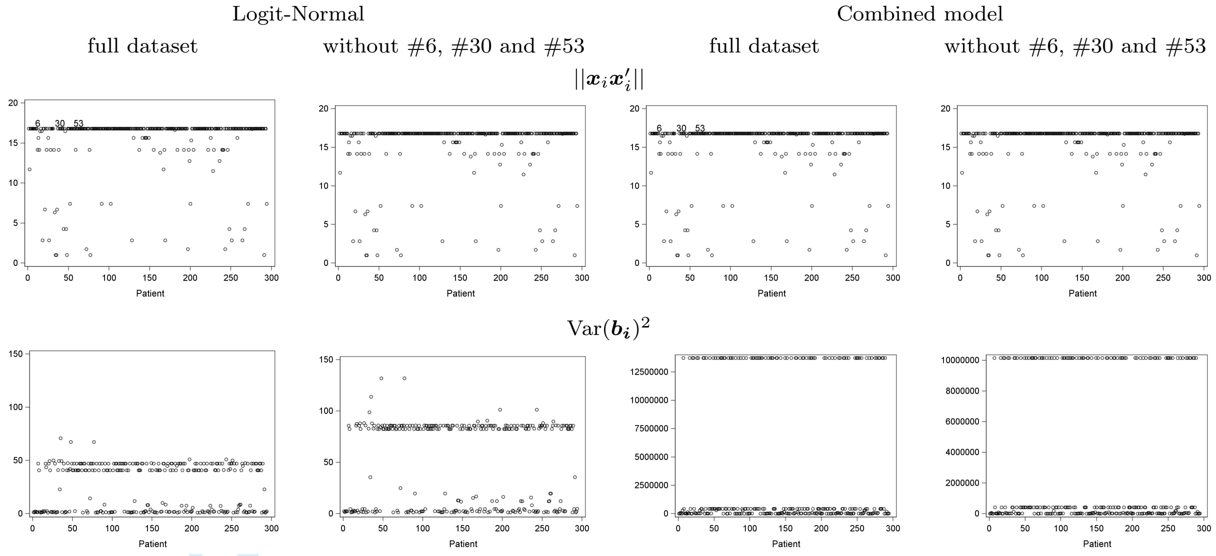

Figure S.8. Onychomycosis Data. Plot of interpretable components of local influence.

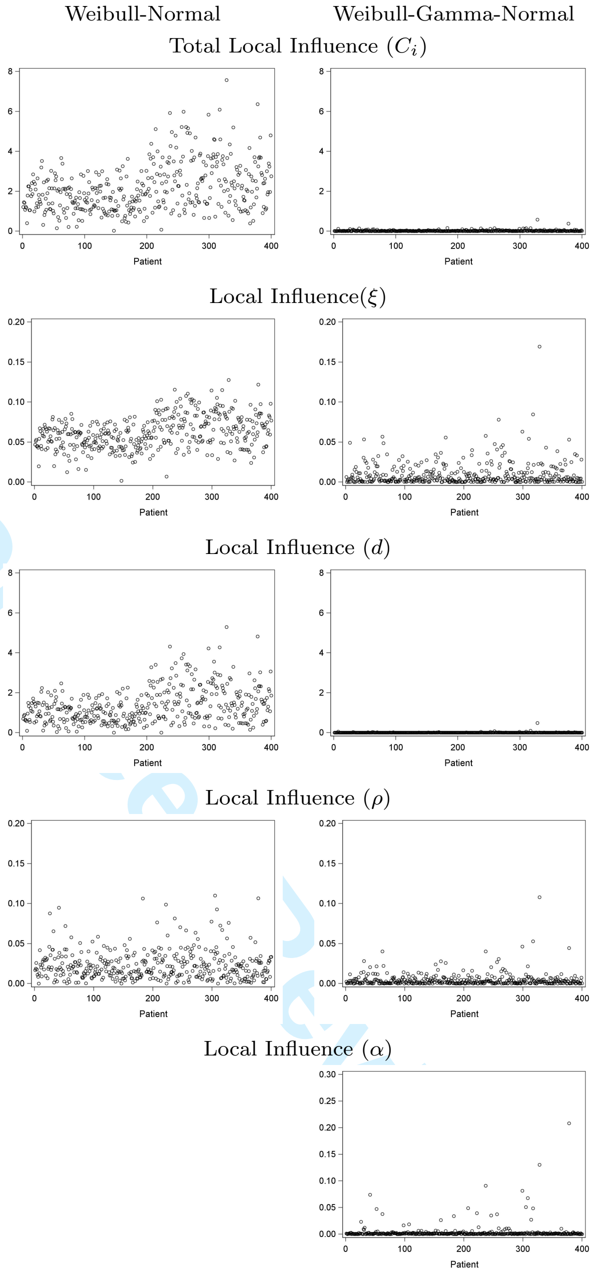

Figure S.9. Recurrent Muscle Soreness Data. Local influence plots.

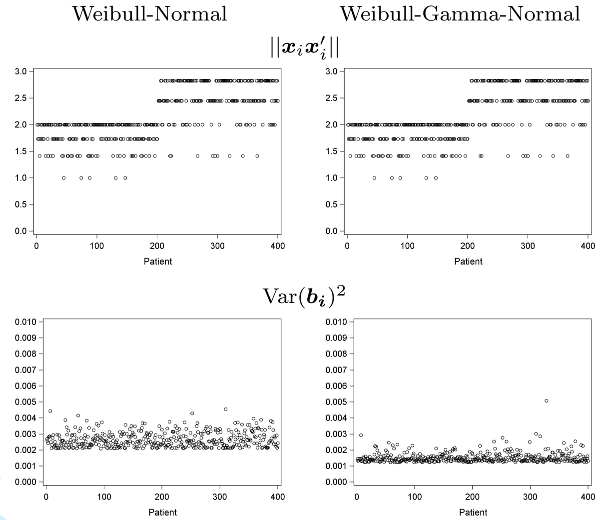

Figure S.10. Recurrent Muscle Soreness Data. Plot of interpretable components of local influence.

S.7. Selected Software Code

S.7.1 *Main Models Code*

```
/**********************************************************************
Software: SAS 9.4
A Clinical Trial in Epileptic Patients (Molenberghs and Verbeke, 2005)
http://www.ibiostat.be/software/default.asp.
**********************************************************************/
***P-N model***;
proc nlmixed data=data2 qpoints=50 hess start;
title Poisson-normal Model;
parms int0=0.5 slope0=-0.1 int1=1 slope1=0.1 sigma=1;
if (trt = 0) then eta = int0 + b + slope0*studyweek;
else if (trt = 1) then eta = int1 + b + slope1*studyweek;
lambda = exp(eta);
model nseizw ~ poisson(lambda);
random b ~ normal(0,sigma**2) subject = idnew ;
predict b out=bi;
predict lambda out=yhat;
estimate "diff in slope" slope1-slope0;
estimate " ratio of slope" slope1/slope0;
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;run;

***PGN model***;
proc nlmixed data=data2 qpoints=50 hess start;
title 'Poisson-combined';
parms int0=0.5 slope0=-0.1 int1=1 slope1=0.1 sigma=1 alpha=5;
if (trt = 0) then eta = int0 + b + slope0*studyweek;
else if (trt = 1) then eta = int1 + b + slope1*studyweek;
lambda = exp(eta);
lambda = exp(eta);
beta=1/alpha; /*hier wordt beta vastgezet*/
loglik=lgamma(alpha+nseizw)-lgamma(alpha)+nseizw*log(beta)-(nseizw+alpha)
 *log(1+beta*lambda) +nseizw*eta;
model nseizw ~ general(loglik);
random b ~ normal(0,sigma**2) subject = idnew ;
predict b out=bi;
predict lambda out=yhat;
estimate "diff in slope" slope1-slope0;
estimate " ratio of slope" slope1/slope0;
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;run;
```

```
/*****************************************************************
Headache Study (McKnight and Van Den Eeden, 1993)
*****************************************************************/
***P-N model***;
proc nlmixed data=data qpoints=50 hess start;
title Poisson-normal Model;
eta = beta0 + b + beta1*(trtn=1);
lambda = d*exp(eta);
model y ~ poisson(lambda);
random b ~ normal(0,sigma**2) subject = id ;
predict b out=bi;
predict lambda out=yhat;
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;
run;

***PGN model***;
proc nlmixed data=data qpoints=50 hess start;
title 'Poisson-combined';
parms beta0=-1.7154  beta1=0.2825  sigma=0.6954  alpha=1;
eta = beta0 + b + beta1*(trtn=1);
lambda = d*exp(eta);
beta=1/alpha; /*hier wordt beta vastgezet*/
loglik=lgamma(alpha+Y)-lgamma(alpha)+Y*log(beta)-(Y+alpha)*log(1+beta*lambda)
       +Y*eta;
model Y ~ general(loglik);
random b ~ normal(0,sigma**2) subject = id ;
predict b out=bi;
predict lambda out=yhat;
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;
run;

/*****************************************************************
A Clinical Trial in Onychomycosis (Molenberghs and Verbeke, 2005)
http://www.ibiostat.be/software/default.asp
*****************************************************************/
***L-N model***;
proc nlmixed data=data2 qpoints=50 hess start;
title 'Logistic-Bernoulli GLMM';
parms int0=-0.7 slope0=-0.22 int1=-0.7 slope1=-0.31 sigma=10;
```

```
if (treatn = 0) then eta = int0 + b + slope0*time;
else if (treatn = 1) then eta = int1 + b + slope1*time;
expeta = exp(eta);
p = expeta /(1 + expeta);
model y ~ binary(p);
random b ~ normal(0,sigma**2) subject = idnew;
predict b out=bi;
estimate "diff in slope" slope1-slope0;
estimate " ratio of slope" slope1/slope0;
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;
run;

***LBN model***;
proc nlmixed data=data2 qpoints=50 hess start;
title 'Overdispersed-Logistic-Bernoulli GLMM' ;
title2 'Random Effect b1 ~ Normal(0,sigma**2)';
title3 'Retriction beta/alpha = const';
parms int0 =-1.54 slope0=-6.49 int1=-16.27 slope1=-8.11 sigma=61 const=0.3 ;
eta = int0*(treatn=0) + slope0*time*(treatn=0)
+ int1*(treatn=1) + slope1*time*(treatn=1) +  b;
expeta = exp(eta);
ll = -log(1+const) + y*eta - y*log(1+expeta)
+ (1-y)*log((1-expeta/(1+expeta)) + const);
model y ~ general(ll);
random b ~ normal(0,sigma**2) subject = idnew;
predict b out=bi;
estimate "diff in slope" slope1-slope0;
estimate " ratio of slope" slope1/slope0;
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;
run;

/*********************************************************************
Recurrent Muscle Soreness (Hosmer and Lemeshow, 1999)
ftp://ftp.wiley.com/public/sci_tech_med/survival/
*********************************************************************/
***W-N model***;
proc nlmixed data=data qpoints=50 hess start;
title 'Weibull model with random-effect';
lambda=1;
eta =  Beta0 + Beta1*(Drug=1) + b;
expeta = exp(eta);
```

```
ll = log(lambda) + log(rho) +  (rho-1)*log(Time) + eta - lambda*(Time**rho)*expeta;
model Time ~ general(ll);
random b ~ normal(0,sigma**2) subject=Patid;
predict b out=bi;
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;
run;

***WGN model***;
proc nlmixed data=data tech=quanew qpoints=50 maxit=1000 hess start;
title 'Weibull model with gamma-normal effects';
bounds alpha > 0;
parms  Beta_0=-3 Beta_1=-0.2  rho=1  sigma=1 alpha=3.3
; /* Beta starting values from Weibull Model via LIFEREG */
lambda=1;
eta =  Beta_0 + Beta_1*(treat=1) + b1;
expeta = exp(eta);
ll = log(lambda) + log(rho) + (alpha+1)*log(alpha)+ (rho-1)*log(Time) + eta
- (alpha+1)*log(lambda*(Time**rho)*expeta + alpha);
model Time ~ general(ll);
random b1 ~ normal(0, sigma**2) subject=id;
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;
run;
```

S.7.2 *Local Influence Code*

```
/*****************************************************************
Software: SAS 9.4
Local Influence for Poisson Normal Models and Extensions
OBJECTIVE: to analyse local influence using numerical approach
 and derived the component of local influence.
DATASET: Epilepsi dataset Molenberghs and Verbeke (2005)
http://www.ibiostat.be/software/default.asp;
VARIABLE DESCRIPTION:
ID id: Patient ID;
study week : time of measurement (repeated)
nseizw: number of seizure per week;
trt: Treatment indicator (1=Drug, 0=Placebo).
DATE: 10/2013 - CENSTAT - Universiteit Hasselt, Diepenbeek, Belgi
AUTHOR : Trias W. Rakhmawati
*****************************************************************/
```

```
options nocenter papersize=A4;

libname data "C:\Users\lucp7702\Documents\PhD\Research\Data\Project 1 -
 LI GLMM\datasets2005";

data data;
set data.Epilepsy;
int=1;
placebo = (trt=0);
treatment = (trt=1);
ptime = placebo*studyweek;
ttime = treatment*studyweek;
run;

*get new id;
 DATA data2;
  SET data;
  by id;
if first.id then idnew+1;
  RUN;

/*calculate the number of repeated measures per subject and the time points*/
proc freq data=data2 noprint;
tables idnew /out=nfrec;
run;
proc sort data=data2;
by idnew;
run;

*use nlmixed the same slope hessian matrix for overall subject;
proc nlmixed data=data2 qpoints=50 hess start;
title Poisson-normal Model;
parms int0=0.5 slope0=-0.1 int1=1 slope1=0.1 sigma=1;
if (trt = 0) then eta = int0 + b + slope0*studyweek;
else if (trt = 1) then eta = int1 + b + slope1*studyweek;
lambda = exp(eta);
model nseizw ~ poisson(lambda);
random b ~ normal(0,sigma**2) subject = idnew ;
predict b out=bi;
predict lambda out=yhat;
estimate "diff in slope" slope1-slope0;
estimate " ratio of slope" slope1/slope0;
```

```
ods output ParameterEstimates = fixedsol;
ods output hessian=hessian;
run;

proc sort data=data2;by idnew;run;

*use nlmixed for each subject;
proc nlmixed data=data2 qpoints=50 maxiter=0 start hess;
title Poisson-normal Model;
parms /data=fixedsol;
if (trt = 0) then eta = int0 + b + slope0*studyweek;
else if (trt = 1) then eta = int1 + b + slope1*studyweek;
lambda = exp(eta);
model nseizw ~ poisson(lambda);
random b ~ normal(0,sigma**2) subject = idnew;
by idnew;
ods output ParameterEstimates = gradientid;
ods output hessian=hessiaanid;
run;

*organizing yhat;
data yhat; set yhat;
keep idnew pred;
run;

/****calculate Local Influence****/

proc iml;
reset print;

/* Matrix with the data to obtain the design matrices for the fixed and
 random effects */
use data2;
labelx = { placebo ptime treatment ttime};
labelz = {int};
labely = {nseizw};
read all var labelx into fixed;
read all var labelz into random;
read all var labely into resp;

p=ncol(fixed);
q=ncol(random);
```

```
/* Matrix with residual */
use yhat;
read all into yhat;
yhat= yhat[,2];

/* Matrix with paramter estimate (fixed n random effect) */
use fixedsol;
read all into fixedsol;
fixedpar= fixedsol[,1];

/* Matrix with 2nd derivative - hessian matrix wtr to parms */
use hessian;
read all into L;
n_L = nrow(L);

L= L[,2:1+n_L];
L_inv=inv(L);

Lb= L[1:n_L-q,1:n_L-q];
Lb_inv=inv(Lb);

Ld= L[p+1:n_L,p+1:n_L];
Ld_inv=inv(Ld);

/* Matrix with individual - 1st derivative */
use gradientid;
read all into Delta;
delta_i= delta[,3];


/* Matrix with the frequencies for each of the subjects */
use nfrec;
read all into nfrec;
id = nfrec[,1];
n_id=nrow(nfrec);

/* Part to calculate the influence measures */
begin = 1;
begin_b = 1;
begin_d = p+1;

do s=1 to n_id ;
end=begin+p+q-1;
```

```
end_b=begin_b+p-1;
end_d=begin_d;

Ci   = 2#delta_i[begin:end, ]`*L_inv*delta_i[begin:end, ];
Ci_b = 2#delta_i[begin_b:end_b, ]`*Lb_inv*delta_i[begin_b:end_b, ];
Ci_d = 2#delta_i[begin_d:end_d, ]`*Ld_inv*delta_i[begin_d:end_d, ];

begin=end+1;
begin_b=end_b+q+1;
begin_d=end_d+p++1;

C_i = C_i//Ci;
Cib = Cib//Ci_b;
Cid = Cid//Ci_d;

index=index//s;
end;

/* Part to calculate the component of influence measures */
begin = 1;

do s=1 to n_id ;
ni = nfrec[s,2];

end=begin+ni-1;

fixedi = fixed[begin:end,];
randomi = random[begin:end,];
respi = resp[begin:end];
yhati = yhat[begin:end];
residi = respi-yhati;

begin = end +1;

rri = sqrt(trace(residi*residi`));
xxi = sqrt(trace(fixedi*fixedi`));

probnorm_rri = probnorm_rri//rri;
probnorm_xxi = probnorm_xxi//xxi;

end;

/*Setting the output dataset with the diagnostic measures */
```

```
out=index||C_i||Cib||Cid||probnorm_rri||probnorm_xxi;
varnames = {'index' 'C_i' 'Ci_b' 'Ci_d' '||rri||' '||xxi||'};

create outdata from out [colname= varnames];
append from out;

close fixedsol;
close hessian;
close nfrec;
close data2;
close gradientid;

quit; /* End of the IML procedure */
```