STATISTICS IN MEDICINE
Statist. Med. 2005 **24**:893–909
Published online 22 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.1974

Influence analysis for linear mixed-effects models

Eugene Demidenko[1,*,†] and Therese A. Stukel[2]

[1] *Dartmouth Medical School, Hanover, NH 03755, U.S.A.*
[2] *Institute for Clinical Evaluative Sciences, G106-2075 Bayview Avenue, Toronto, Ontario, Canada M4N 3M5*

SUMMARY

In this paper, we extend several regression diagnostic techniques commonly used in linear regression, such as leverage, infinitesimal influence, case deletion diagnostics, Cook's distance, and local influence to the linear mixed-effects model. In each case, the proposed new measure has a direct interpretation in terms of the effects on a parameter of interest, and collapses to the familiar linear regression measure when there are no random effects. The new measures are explicitly defined functions and do not necessitate re-estimation of the model, especially for cluster deletion diagnostics. The basis for both the cluster deletion diagnostics and Cook's distance is a generalization of Miller's simple update formula for case deletion for linear models. Pregibon's infinitesimal case deletion diagnostics is adapted to the linear mixed-effects model. A simple compact matrix formula is derived to assess the local influence of the fixed-effects regression coefficients. Finally, a link between the local influence approach and Cook's distance is established. These influence measures are applied to an analysis of 5-year Medicare reimbursements to colon cancer patients to identify the most influential observations and their effects on the fixed-effects coefficients. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: case deletion; infinitesimal influence; local influence; random effects; repeated measurements; sensitivity analysis

1. INTRODUCTION

Statistical influence analysis is defined as the assessment of the effects of small perturbations in the data on the outcomes of the analysis [1]. The general aim is to assess the robustness of the model to the observed data points. Several types of influence analyses have been developed for linear regression models [2]. First, standardized residuals are typically used to detect outliers; however, the presence of such outliers does not necessarily affect the model fit or the statistical inference. Second, leverage, defined as the identification of data

*Correspondence to: Eugene Demidenko, HB 7927, Section of Biostatistics and Epidemiology, Dartmouth Medical School, Hanover, NH 03755, U.S.A.
†E-mail: eugene.demidenko@dartmouth.edu

Contract/grant sponsor: NCI; contract/grant number: CA 52192

Received June 2003
Accepted June 2004

points that influence the fitted values, are detected by exploring large values of the diagonal elements of the projection (or 'hat') matrix [1]. Rao and Toutenburg [3] noted that leverage is identical to the partial derivative of the predicted value with respect to the corresponding dependent variable. Laurent and Cook [4, 5] adapted this approach to non-linear regression and introduced Jacobian leverage as the partial derivative of the predicted values with respect to the dependent variable. Third, case deletion diagnostics determine how deletion of a specific observation affects the estimated regression coefficients. In linear regression, these can be obtained without recomputing the ordinary least squares (OLS) estimate since a closed form solution to the updated OLS estimate exists [6, 7]. Based on this formula, Cook [8] and Cook and Weisberg [1] suggested analysing the standardized squared distance between the original OLS estimate and the estimate after case deletion. This became known as Cook's distance and is computed in commonly used statistical packages such as SAS and S-plus. Fourth, infinitesimal (I) influence assesses how a small (infinitesimal) change in the value of the observation or its weight in the estimating equation affects characteristics of interest, such as predicted values, regression coefficients, or residuals. I-influence collapses to the evaluation of the partial derivative of the characteristic with respect to the observation or its weight. In a seminal paper, Pregibon [9] applied the infinitesimal influence approach to logistic regression. Finally, Cook [10] generalized infinitesimal influence to investigate the effects of individual observations on the log-likelihood function (or 'log-likelihood displacement'); this type of influence analysis was called local influence. Lesaffre and Verbeke [11] applied the local influence approach to linear mixed-effects models, illustrating its use in various examples [12]. Pan and Fang [13] provide a detailed discussion of the local influence approach and Cook's distance for balanced growth curve models. Frans *et al.* [14] introduce the conditional Cook's distance and decompose it into three parts. A general discussion of influence analysis and model diagnostics can be found in a recent book by Demidenko [15].

The goal of this paper is to generalize several common measures of influence for the fixed-effects parameters of the linear mixed-effects (LME) model. We focus on simple, easily computable approximations that are particularly useful for large data sets, where cluster deletion influence cannot be practically implemented by iteratively removing one cluster and re-estimating. An advantage of the simple influential measures is that they permit a visual analysis crucial in diagnostics analysis, through simple plots. In Section 2, we generalize the leverage measure to a leverage matrix for the LME model. In Section 3, we adapt the principles of I-influence analysis to assess the effects of small changes in the dependent or independent variables on the estimates of the fixed-effects parameters. In Section 4, we propose a generalization of Cook's distance to the LME model. In Section 5, we derive cluster deletion diagnostics in the situation when the variance of the random effects is unknown. In Section 6, we provide an explicit matrix formula to compute local influence for the fixed-effects parameters. In the last section, we apply the new influence measures to a study of Medicare reimbursement data.

2. LEVERAGE FOR THE LINEAR MIXED-EFFECTS MODEL

We consider the linear mixed-effects model as defined by Laird and Ware [16]

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, N \tag{1}$$

where $\mathbf{y}_i$ is the $n_i \times 1$ vector of observations of the dependent variable on the ith cluster, $\mathbf{X}_i$ is the $n_i \times m$ matrix of fixed-effects covariates, $\boldsymbol{\beta}$ is the $m \times 1$ vector of fixed-effects parameters, $\mathbf{Z}_i$ is the $n_i \times k$ matrix of random effects, $\mathbf{b}_i$ is the $k \times 1$ vector of random effects parameters with zero mean and covariance matrix $\sigma^2\mathbf{D}$, and $\boldsymbol{\varepsilon}_i$ is the $n_i \times 1$ vector of errors with zero mean and covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}_i) = \sigma^2\mathbf{R}_i(\boldsymbol{\gamma})$. There are N clusters and n_i observations within the ith cluster. The form of the $n_i \times n_i$ matrix function $\mathbf{R}_i$ is assumed known but the parameter vector $\boldsymbol{\gamma}$ is unknown and estimated using unrestricted or restricted maximum likelihood. These assumptions imply that the covariance matrix of $\mathbf{y}_i$ is $\sigma^2\mathbf{V}_i$, where $\mathbf{V}_i = \mathbf{R}_i(\boldsymbol{\gamma}) + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'$. When components of $\boldsymbol{\varepsilon}_i$ are independent, $\mathbf{V}_i = \mathbf{I} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'$. When the components follow a first-order autoregressive AR(1) form, $(R_i)_{jk} = \gamma^{|j-k|}$. It is assumed that $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are normally and independently distributed. If $\boldsymbol{\gamma}$ and $\mathbf{D}$ were known, the estimates of the fixed-effects parameters can be found using generalized least squares (GLS) as $\widehat{\boldsymbol{\beta}} = \mathbf{M}^{-1}\mathbf{s}$, where

$$\mathbf{M} = \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i, \quad \mathbf{s} = \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i \tag{2}$$

However, the LME model is typically estimated using restricted maximum likelihood (REML), which simultaneously produces an estimate of $\mathbf{D}$ and $\boldsymbol{\beta}$ [17, 18, 15].

Leverage is the partial derivative of the predicted value with respect to the corresponding dependent variable. Hence, the ith leverage indicates how the predicted value of the ith case is influenced by the ith observation. For the standard linear regression model, $y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$, $i = 1, \ldots, N$, it can be shown that leverage $h_i = \partial\widehat{y}_i/\partial y_i = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$ where $\widehat{y}_i = \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$. The matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is often referred to as the *hat* matrix. The h_i are positive and sum to m, the number of estimated parameters so that the 'expected' leverage is m/N [19]. Large values of h_i point to influential observations. This interpretation is important for the generalization to the linear mixed-effects (LME) model. Analogously to linear regression, we define the $n_i \times n_i$ leverage matrix for the LME model as

$$\mathbf{H}_i = \frac{\partial\widehat{\mathbf{y}}_i}{\partial\mathbf{y}_i} \tag{3}$$

where $\widehat{\mathbf{y}}_i = \mathbf{X}_i\widehat{\boldsymbol{\beta}} + \mathbf{Z}_i\widehat{\mathbf{b}}_i$ is the predicted outcome, conditional on the estimated random effect, $\widehat{\mathbf{b}}_i$. Matrix (3) collapses to the usual scalar h_i when each cluster contains only one observation and the variance of the random effects is zero. After differentiation and assuming $\mathbf{V}_i$ is fixed, the leverage matrix can be represented as the sum of two matrices, $\mathbf{H}_i = \mathbf{H}_{i1} + \mathbf{H}_{i2}$, where

$$\mathbf{H}_{i1} = \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}, \quad \mathbf{H}_{i2} = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'\mathbf{V}_i^{-1}(\mathbf{I} - \mathbf{H}_{i1}), \quad i = 1, \ldots, N \tag{4}$$

$\mathbf{H}_{i1} = \mathbf{X}_i\widehat{\boldsymbol{\beta}}$ is the leverage for the fixed effects, or the marginal leverage. $\mathbf{H}_{i2}$ is the leverage associated with the random effects, which vanishes when $\mathbf{D} = \mathbf{0}$. The sum of the traces of the fixed-effects leverage matrix is equal to the dimension of $\boldsymbol{\beta}$, similarly to linear regression,

$$\begin{aligned}\mathrm{tr}\left(\sum_{i=1}^{N}\mathbf{H}_{i1}\right) &= \sum_{i=1}^{N}\mathrm{tr}\,\mathbf{H}_{i1} = \sum_{i=1}^{N}\mathrm{tr}(\mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i) \\ &= \mathrm{tr}\left(\mathbf{M}^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right) = \mathrm{tr}\,\mathbf{I}_m = m \end{aligned} \tag{5}$$

Thus, the average value of $\mathrm{tr}\,\mathbf{H}_{i1}$ is m/N.

The proposed leverage measure is $\operatorname{tr}\mathbf{H}_i$. A larger than average value of $\operatorname{tr}\mathbf{H}_{i1}$ points to an influential cluster i due to an outlier in the **X**-space and for $\operatorname{tr}\mathbf{H}_{i2}$, to an outlier in the **Z**-space. Leverage is affected by cluster size. One could use the average leverage $n_i^{-1}\operatorname{tr}\mathbf{H}_i$ to adjust for cluster size, but decomposition (5) would not hold. After an influential cluster is identified, one may further examine $\mathbf{H}_i$ for individual observations within cluster i having high leverage. With the usual linear model, leverage analysis detects outliers in the covariate space since the expression does not involve the dependent observations. For linear mixed-effects models, both the covariates and the random effects may influence the leverage measure.

3. INFINITESIMAL INFLUENCE FOR THE FIXED-EFFECTS COEFFICIENTS

We next determine if the estimates of the fixed-effects coefficients are sensitive to small changes in the dependent or independent variables, using the infinitesimal approach proposed by Pregibon [9]. One may use the infinitesimal influence approach to assess how small perturbations in the dependent variable $\mathbf{y}_i$ affect $\widehat{\boldsymbol{\beta}}$, by computing

$$\frac{\partial\widehat{\boldsymbol{\beta}}}{\partial\mathbf{y}_i}=\mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1} \tag{6}$$

The (k,l)th element of matrix (6) can be interpreted as the degree of change in the kth coefficient per unit change in y_{il}. One may also compute the influence of independent variables using

$$\begin{aligned}\frac{\partial\widehat{\boldsymbol{\beta}}}{\partial x_{ijk}}&=-\mathbf{M}^{-1}\left(\frac{\partial\mathbf{M}}{\partial x_{ijk}}\right)\mathbf{M}^{-1}\mathbf{s}+\mathbf{M}^{-1}\frac{\partial\mathbf{s}}{\partial x_{ijk}}\\&=\mathbf{M}^{-1}(\mathbf{E}_{ijk}'\mathbf{V}_i^{-1}\mathbf{y}_i-[\mathbf{E}_{ijk}'\mathbf{V}_i^{-1}\mathbf{X}_i+\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{E}_{ijk}]\widehat{\boldsymbol{\beta}})\\&=\mathbf{M}^{-1}(\mathbf{E}_{ijk}'\mathbf{V}_i^{-1}\mathbf{r}_i-\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{E}_{ijk}\widehat{\boldsymbol{\beta}}),\quad j=1,\dots,n_i,\ \ k=1,\dots,m\end{aligned} \tag{7}$$

where $\mathbf{E}_{ijk}$ is an $n_i\times m$ matrix of zeros except for the (j,k)th element which is equal to 1, x_{ijk} is the (j,k)th element of the matrix $\mathbf{X}_i$ and $\mathbf{r}_i=\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}}$ is the $n_i\times 1$ residual vector. These formulae collapse to those in Reference [1] when there are no random effects and a single observation per cluster.

The lth element of vector (7) has a clear interpretation as the change in the lth coefficient of the fixed-effects parameters with respect to a unit change in x_{ijk}. The influence of a particular covariate x_{ijk} on the vector of parameters $\widehat{\boldsymbol{\beta}}$ is easily assessed by plotting the elements of (7) against the cluster index i separately for each coefficient β_l.

4. CLUSTER DELETION DIAGNOSTICS AND COOK'S DISTANCE: THE CASE OF KNOWN D

We next discuss cluster deletion diagnostics for the linear mixed-effects model when **D** is known. Cluster deletion diagnostics with unknown **D** are discussed in the next section. For the linear regression model,

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)} = \frac{r_i}{1 - h_i} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i' \tag{8}$$

where $\widehat{\boldsymbol{\beta}}$ is the $m \times 1$ OLS estimate based on the entire sample, $\widehat{\boldsymbol{\beta}}_{(i)}$ is the OLS estimate with the ith case deleted, $\mathbf{x}_i$ is the ith row-vector of $\mathbf{X}$, r_i is the residual, and h_i is the leverage [6, 7]. It is shown in the appendix that for the LME model (1) with known $\mathbf{D}$, this takes the form

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)} = \mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{I} - \mathbf{H}_{i1})^{-1}\mathbf{r}_i \tag{9}$$

where matrices $\mathbf{M}$ and $\mathbf{H}_{i1}$ are defined in (2) and (4). This formula permits a simple calculation of the cluster deletion effect without re-estimating the model N times. It also gives rise to the following generalization of Cook's distance for the linear mixed-effects model (1). A similar formula for case deletion was derived in Reference [20] in the framework of RML estimation, and in a case of cluster deletion in Reference [21].

For the linear regression model, Cook [8] proposed that the influence of the ith case be measured by the squared distance between $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{(i)}$ scaled by the variance, which is expressed as

$$M_i^2 = \frac{1}{m\widehat{\sigma}^2} (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})'\mathbf{M}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}) = \frac{1}{m\widehat{\sigma}^2} \frac{r_i^2}{(1 - h_i)^2} \mathbf{x}_i\mathbf{M}^{-1}\mathbf{x}_i' \tag{10}$$

using (8). Analogously, using (9), Cook's distance for the linear mixed-effects model takes the form

$$M_i^2 = \frac{1}{m\widehat{\sigma}^2} \mathbf{r}_i'(\mathbf{I} - \mathbf{H}_{i1})^{-1}\mathbf{V}_i^{-1}\mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{I} - \mathbf{H}_{i1})^{-1}\mathbf{r}_i \tag{11}$$

For balanced data ($n_i = n$, $\mathbf{X}_i = \mathbf{X}$, $\mathbf{Z}_i = \mathbf{Z}$), this measure is presented in Reference [21] and a recent book [13, Section 4.3.1]. Formula (11) is more general, however, because it covers the non-balanced case as well.

The relative influence of various clusters i is judged by the corresponding values of M_i. Again, (11) collapses to Cook's distance (10) for the linear regression model with no random effects and a single observation per cluster. Plotting Cook's distance against the cluster index i highlights influential clusters; however, since Cook's distance measures the combined effects of $\mathbf{r}_i$ and $\mathbf{H}_i$, it does not distinguish among clusters that have high leverage and those that are outliers.

5. INFINITESIMAL CLUSTER DELETION DIAGNOSTICS: THE CASE OF UNKNOWN D

We next discuss cluster deletion diagnostics in the situation of unknown $\mathbf{D}$. When $\mathbf{D}$ is unknown, there does not exist a closed-form update formula for $\boldsymbol{\beta}$ when the ith cluster is deleted. We therefore adapt infinitesimal case deletion techniques [9] to cluster deletion for linear mixed-effects models (1) since we are interested in how deletion of the ith cluster influences $\widehat{\boldsymbol{\beta}}$. Following Pregibon's notation, the joint log-likelihood function and the score equations, in general terms, are represented as

$$l = \sum_{j \neq i} l_j + w l_i \quad \text{and} \quad \sum_{j \neq i} \mathbf{S}_j + w\mathbf{S}_i = \mathbf{0} \tag{12}$$

respectively, where w is viewed as a control variable and $\mathbf{S}$ is the derivative of l with respect to unknown parameters. Then the sensitivity of the MLE $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(w)$ to deletion of the ith log-likelihood contribution is assessed using the derivative as

$$\left. \frac{\mathrm{d}\widehat{\boldsymbol{\theta}}(w)}{\mathrm{d}w} \right|_{w=1} = -\left(\sum_{j=1}^{N} \frac{\partial \mathbf{S}_j}{\partial \boldsymbol{\theta}} \right)^{-1} \mathbf{S}_i \tag{13}$$

where the right-hand side is evaluated at the MLE, $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}} = \widehat{\boldsymbol{\theta}}(1)$. The inverse matrix is the asymptotic covariance matrix and the matrix under inverse sign is the Hessian of the complete log-likelihood. Thus the magnitude of the infinitesimal cluster deletion diagnostics is determined by the contribution of the ith score equation. This formula is general and can be applied to the infinitesimal cluster deletion of linear or non-linear mixed models. Note, as follows from (13), infinitesimal deletion reduces to the computation of the Hessian.

The MLE (or RML) $\widehat{\boldsymbol{\beta}}$ is then a function of w, and its sensitivity to the deletion of the ith cluster can be measured using the derivative $\mathrm{d}\widehat{\boldsymbol{\beta}}(w)/\mathrm{d}w$ evaluated at $w = 1$. We show in the appendix that the asymptotic approximation to case deletion diagnostics in the LME model when $\mathbf{D}$ is unknown is

$$\left. \frac{\mathrm{d}\widehat{\boldsymbol{\beta}}(w)}{\mathrm{d}w} \right|_{w=1} \simeq \mathbf{M}^{-1} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{r}_i \tag{14}$$

which should work well in large samples. This derivative is estimated by replacing $\boldsymbol{\beta}$ and $\mathbf{D}$ by their ML or RML estimates. By asymptotics, we refer to the situation when the number of clusters (N) goes to infinity but the number of observations per cluster (n_i) is bounded. Since asymptotically, $\widehat{\boldsymbol{\beta}}$ and the variance parameters are independent, the information and the inverse matrix in (13) are block-diagonal and therefore the infinitesimal cluster deletion diagnostics can be written as separate terms. Previous work has shown that substitution of any consistent estimator of $\mathbf{D}$ in the generalized least squares estimator leads to a consistent, asymptotically normally distributed, and efficient estimate of $\boldsymbol{\beta}$ that is equivalent to the MLE for linear mixed-effects models [22]. Thus, substituting either the MLE or the RML of $\mathbf{D}$ into (14) will produce consistent estimates of the influence measures. Finally, note the similarity of this formula to the previous formula (9) when $\mathbf{D}$ is known.

As mentioned above, to obtain the derivative of the estimate with respect to the cluster weight for small N, one needs to compute the Hessian of the log-likelihood. Sensitivity to cluster deletion may be studied for coefficients or variance parameters. Relevant computations of the Hessian in the framework of the LME model (1) were presented elsewhere [23, 15]. For the special case of the random intercept model, formulae are given in Section 7.

6. LOCAL INFLUENCE

Using the local influence approach, we seek to determine how the log-likelihood function changes with respect to various characteristics of interest such as values of the dependent or independent variable, or case deletion. Beckman *et al.* [24] applied the local influence approach to the random effects model. Lesaffre and Verbeke [11] and Verbeke and Molenberghs [12] applied the infinitesimal cluster deletion diagnostics based on the idea of Pregibon [9] to the linear mixed-effects model.

The local influence approach can be viewed as infinitesimal case deletion diagnostics in the general maximum likelihood framework, and is summarized as follows. Let $\boldsymbol{\theta}$ be the vector of unknown parameters estimated by maximizing the log-likelihood function $l=\sum_{i=1}^{N} l_i(\boldsymbol{\theta})$. Let $\widehat{\boldsymbol{\theta}}$ be the MLE and $\widehat{\boldsymbol{\theta}}_{(i)}$, the MLE after assigning weight w to the ith case in the log-likelihood function, $l(w)$. Cook [10] suggested measuring the influence of the ith observation on the local behaviour of the log-likelihood displacement, $2(l-l(w))$, defined in terms of its normal curvature, for small perturbations of w from 1. Evaluation at $w=1$ describes local changes in $l(w)$ at the MLE. Since this normal curvature influence measure is a non-linear function of w, he suggested using a first-order approximation, $C_i=\boldsymbol{\Delta}_i{}'\mathscr{I}^{-1}\boldsymbol{\Delta}_i$, where $\boldsymbol{\Delta}_i=\partial l_i/\partial\boldsymbol{\theta}$ is the derivative of the ith log-likelihood function and $\mathscr{I}$ is the Fisher information matrix evaluated at the MLE. In linear mixed-effects models, estimates for the fixed effects and the variance parameters are independent since the information matrix has a block diagonal form. Thus, C_i may be decomposed as $C_i=C_{i,\boldsymbol{\beta}}+C_{i,\boldsymbol{\alpha}}$ where $\boldsymbol{\alpha}=(\sigma^2,\mathrm{vech}(\mathbf{D}))'$ is the $1+k(k+1)/2$ vector of variance parameters [11, 12]. To compute C_i we need the first and second derivatives of the log-likelihood function. Lesaffre and Verbeke provided these derivatives in co-ordinate form which is cumbersome. Focusing on $\boldsymbol{\beta}$, we write $C_{i,\boldsymbol{\beta}}$ as a compact closed-form matrix formula

$$C_{i,\boldsymbol{\beta}}=\widehat{\sigma}^{-2}\mathbf{r}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{r}_i \tag{15}$$

where $\widehat{\sigma}^2$ is the MLE estimate, $\mathbf{r}_i=\mathbf{y}_i-\widehat{\mathbf{y}}_i$ is the residual vector for the ith cluster, and $\mathbf{M}$ and $\mathbf{V}_i$ are defined earlier. The derivation of (15) is presented in the appendix. Cook's distance is closely related to the local influence function of $\boldsymbol{\beta}$. In fact, it is easy to see that (15) and (11) coincide up to a factor of $1/m$ when the matrix $\mathbf{H}_{i1}=\mathbf{0}$. Finally $C_{i,\boldsymbol{\beta}}$ measures the normal curvature of the log-likelihood function, and is therefore is less directly interpretable than the other proposed measures.

7. FACTORS AFFECTING TOTAL MEDICARE REIMBURSEMENTS FOR COLON CANCER PATIENTS

Fisher *et al.* [25] reported aggregate Medicare spending on a cohort of $N=10\,109$ colon cancer patients up to 5 years after initial hospitalization for potential curative resection for colorectal cancer. Patients in the study were initially admitted for colon cancer surgery between 1993 and 1995, were eligible for both Medicare part A (hospital) and B (physician) services, and were not enrolled in an HMO. Characteristics of the patients as well as reimbursements for all hospital and physician services were obtained from Medicare claims data. Patients were followed through December 31, 1997, censoring for mortality. They developed a measure of regional intensity of medical services consisting of total Medicare spending to resident

Table I. Results of estimation of model (16) using REML.

	Independent errors			AR(1), $\hat{\gamma} = 0.12$		
Variable	Coeff	SE	p-value	Coeff	SE	p-value
Intercept	6.72	0.103	<0.0001	6.74	0.102	<0.0001
Male	0.053	0.017	0.001	0.053	0.016	0.001
White race	0.074	0.032	0.0208	0.074	0.0317	0.0191
Stage II	0.243	0.021	<0.0001	0.236	0.021	<0.0001
Stage III	0.328	0.023	<0.0001	0.319	0.029	<0.0001
Age (decade)	-0.049	0.012	<0.0001	-0.047	0.012	<0.0001
Charlson index	0.146	0.012	<0.0001	0.144	0.011	<0.0001
HRR intensity (per \$1000)	0.076	0.004	<0.0001	0.075	0.004	<0.0001
Interval 1 (0–6 months)	2.56	0.016	<0.0001	2.55	0.016	<0.0001
Time (6-month intervals)	-0.114	0.003	<0.0001	-0.121	0.003	<0.0001
$\hat{\sigma}_b$	0.59			0.53		
$\hat{\sigma}$	1.15			1.17		

patients in their last 6 months of life. Overall spending on patients is strongly related to this regional intensity measure, independent of patient illness. We provide further analyses of the relationships between spending and patient illness and regional intensity of medical services. For each patient, we grouped total Medicare reimbursements in 6 month intervals post-colectomy. We used a log transformation to reduce the right skewing of the data, and we added \$100 to all non-missing reimbursements to attenuate the effects of the compound distribution created by the large proportion of zero charges, since between 5 and 29 per cent of patients had no medical expenses in a specific 6 month interval.

The following random intercept model is a special case of linear mixed-effects model (1) and was used to analyse these data since we anticipate individual patients with unusually large or small expenditures.

$$\ln(y_{it} + 100) = \beta_0 + b_i + \beta_1 x_{it} + \cdots + \beta_m x_{mt} + \varepsilon_{it}, \quad t = 1, \ldots, 10, \;\; i = 1, \ldots, N \tag{16}$$

where y_{it} represents the logarithm of Medicare reimbursements for patient i in each 6-month period t ($t = 1, \ldots, 10$), β_0 is the population-averaged intercept, b_i is the patient random effect, and x_i are covariates. It is assumed that b_i and ε_{it} are independent, normally distributed random variables with zero means and variances, σ_b^2 and σ^2, respectively. Under independent error assumption, $\mathbf{V}_i = \sigma^2(\mathbf{I} + d\mathbf{1}_i\mathbf{1}_i')$ where $d = \sigma_b^2/\sigma^2$ and $\mathbf{1}_i$ is the $n_i \times 1$ vector of ones. To address serial correlation we use the autoregression model of the first order, AR(1). Patient covariates include age, gender, race (white/black), stage (I, II, III+), Charlson comorbidity score and regional medical intensity. Time-dependent covariates include a continuous time term and an indicator variable for first 6 months of follow-up. The random intercept model assumes that there is additional variation in spending across patients that is not accounted for by the patient illness factors.

The results of restricted maximum likelihood estimation using the function `lme` in S-plus are presented in Table I. Since the autocorrelation coefficient is fairly small we use independent assumption for the influence analyses in this example (also autoregression would considerably

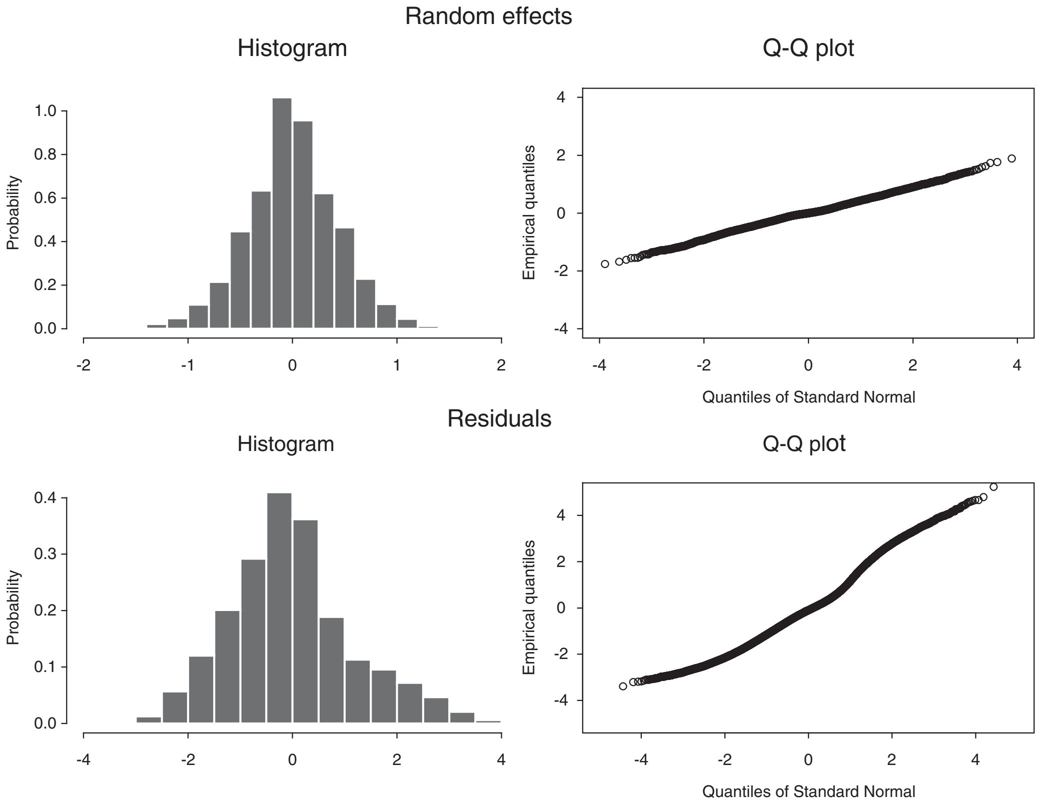

Figure 1. Histograms and Q–Q plots of the patient random effects and the residuals for the linear mixed-effects model.

increase the time of computation). There is a large peak in medical spending during the first 6 months after colectomy due to the high costs associated with the initial hospital admission, colon cancer treatment and surveillance; after this, spending decreases by 10.8 per cent (10.2, 12.0 per cent) per 6-month interval. On average, spending is 5 per cent (2, 9 per cent) higher for males, 8 per cent (1, 15 per cent) higher for whites and 39 per cent (33, 45 per cent) higher for stage III patients. Spending on an individual patient increases by 7.9 per cent (7.0, 8.4 per cent) per \$1000 increase in the regional intensity measure independent of patient illness factors. It decreases with age and increases with number and severity of comorbidities.

The histogram and Q–Q plots for the random effects and the residuals, respectively, are plotted in Figure 1; the normal distribution assumptions appear to hold with slight remaining skewing of the residuals and short tails for the patient random effects. Variation in reimbursements among patients is greater than predicted by the demographic and illness factors since the intraclass correlation coefficient is $r = 0.21$. For the random intercept model, $\text{tr}(\mathbf{H}_{i2}) = n_i d/(1 + n_i d)$ where $d = \sigma_b^2/\sigma^2$, depends only on n_i, and is therefore relatively uninformative.

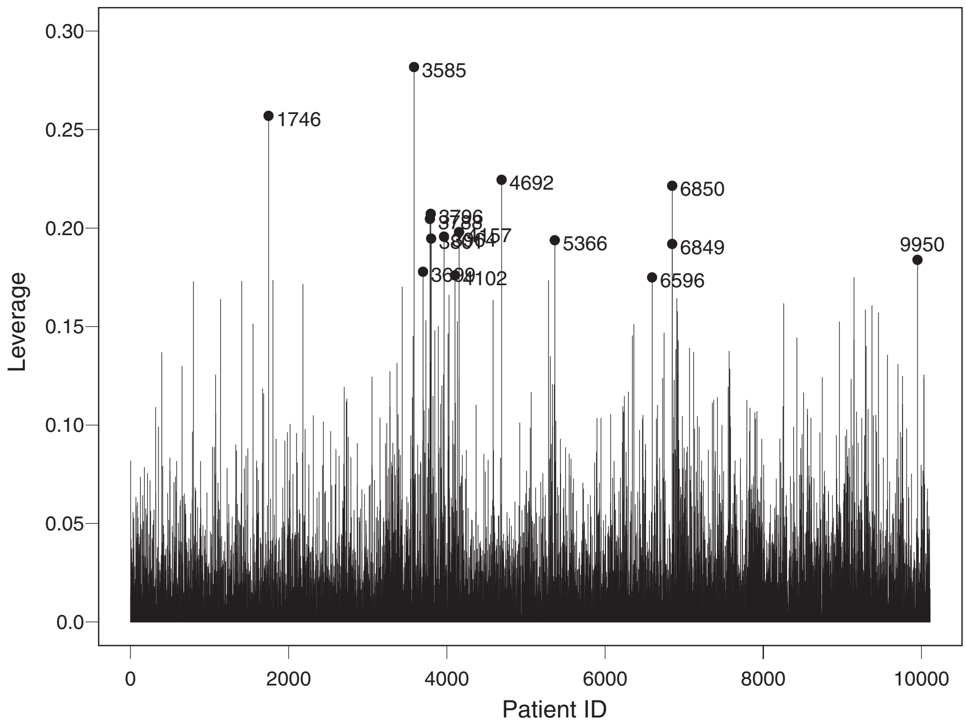

Figure 2. Plot of leverage as $\mathrm{tr}(\mathbf{H}_{i1})$ against patient ID, identifying the 15 most influential patients.

These Medicare data are an example where direct cluster (patient) deletion diagnostics would be prohibitive because it took about 30 s to run one lme model using S-plus 6.1 on a 3 GHz PC with 1 GB RAM. With more than 10 000 clusters, the full cluster deletion diagnostics would take 83 h. Although this is a conservative estimate, because one could start Newton–Raphson iterations from the MLE, complete re-estimation of the model might be computationally intensive for large data sets such as used in health services research studies. The idea behind our model diagnostics is to derive a simple, easily computable approximation to serve as an explanatory tool to test several competing models.

To identify subjects who influence the regression estimates, we focus on the fixed-effects leverage based on $\mathbf{H}_{i1}$. Figure 2 is a plot of the leverage $\mathrm{tr}\,\mathbf{H}_{i1}$ against the cluster index i, where $i = 1, \ldots, N$. All but two of the 15 most influential patients were black and had much lower than average costs. Re-running the model after deletion of these 15 patients decreased the coefficient of race by about half to 0.043 (S.E. $= 0.032$), making it non-significant; other coefficients changed negligibly. Figure 3 displays the infinitesimal influence of regional medical intensity on the fixed-effects parameters using (7), reported as a relative difference. The coefficient for race is again the most sensitive to changes in regional intensity. The 15 patients with the most influence on the race coefficient were all black and most had higher than average costs but lived in average intensity regions. Re-running the model after deletion of these 15 patients increased the coefficient of race to 0.113 (S.E. $= 0.032$), implying that spending for whites is 12 per cent (5, 19 per cent) higher than for blacks; other coefficients

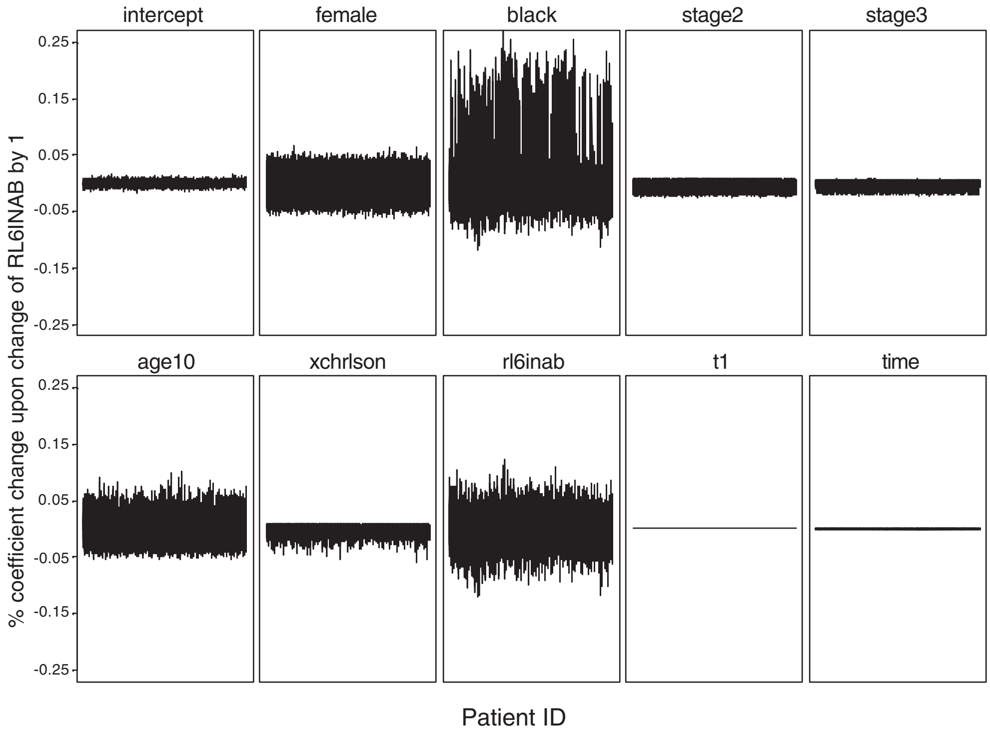

Figure 3. Infinitesimal influence of regional medical intensity on each fixed-effects coefficient.

changed negligibly. Removing both sets of outliers simultaneously produced a regression coefficient for race of 0.082 (S.E. $= 0.033$), which is only slightly higher than the original race effect.

Figure 4 displays the relative change in parameter values after deletion of individual patients using update formula (9); again, the coefficient for race is the most sensitive to individual patient values. In Figure 5, we plot Cook's distance M_i^2 against patient index i using formula (11). The 15 most influential patients almost all lived in regions with high intensity medical services. Figure 6 shows the density of the Cook's distance values. This influence measure has a long right tail indicating the presence of several strong outliers.

Re-running the model after deletion of these 15 patients changed the coefficient of race negligibly to 0.085 (S.E. $= 0.032$). For this model, Cook's distance (11), infinitesimal cluster deletion diagnostics (14), and the local influence approach (15) all produced the identical set of influential observations because (i) N is large and $\mathbf{H}_i = O(1/N)$ and (ii) there are no extreme observations in the covariate space. We therefore do not report local influence.

Both the leverage and the infinitesimal influence analyses were able to identify 15 patients with much lower or higher than average costs who influenced the estimated effect of race on total Medicare expenditures; however, further analysis demonstrated that these sets of influential clusters were balancing each other's effects.

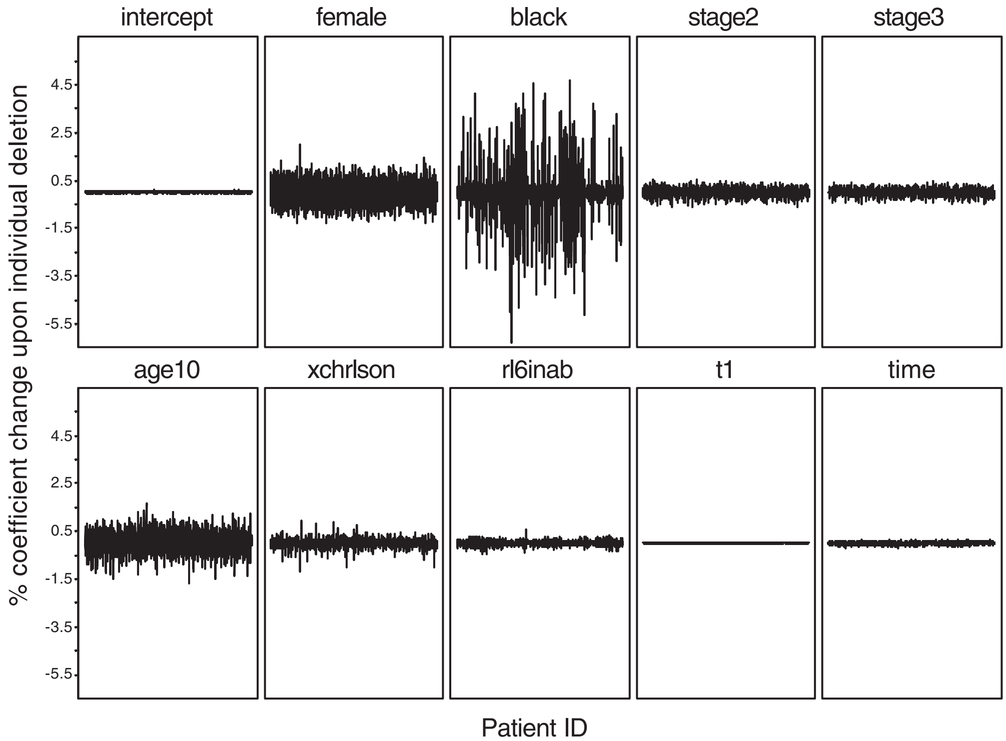

Figure 4. Relative change in each fixed-effects coefficient after individual patient deletion.

8. DISCUSSION

Regression diagnostics are an important component of statistical inference for detecting influential observations or clusters. These diagnostics are especially important for mixed-effects models where parameter estimates cannot be obtained in closed form solution because estimation requires computing the partial derivative of an implicitly defined function. The proposed infinitesimal influence measures are explicitly defined formulas and do not necessitate a maybe time consuming re-estimation of the regression coefficient $\widehat{\boldsymbol{\beta}}$, especially for cluster deletion diagnostics such as (9), (11) and (14).

For some measures such as (9), a closed-form expression does not exist if the matrix $\mathbf{D}$ is unknown. For this case, we provide the large sample approximation (14) based on the concept of infinitesimal cluster deletion. We demonstrate that the two approaches yield similar formulae. Since substitution of any consistent estimator of $\mathbf{D}$ in the generalized least squares estimations leads to a consistent, asymptotically normally distributed, and efficient estimate of $\boldsymbol{\beta}$ that is equivalent to the MLE for linear mixed-effects models [22], substituting either the MLE or the RML of $\mathbf{D}$ will produce consistent estimates of the influence measures. Moreover, since the distribution of $\widehat{\boldsymbol{\beta}}$ is independent of the particular consistent estimator of $\mathbf{D}$ used, one

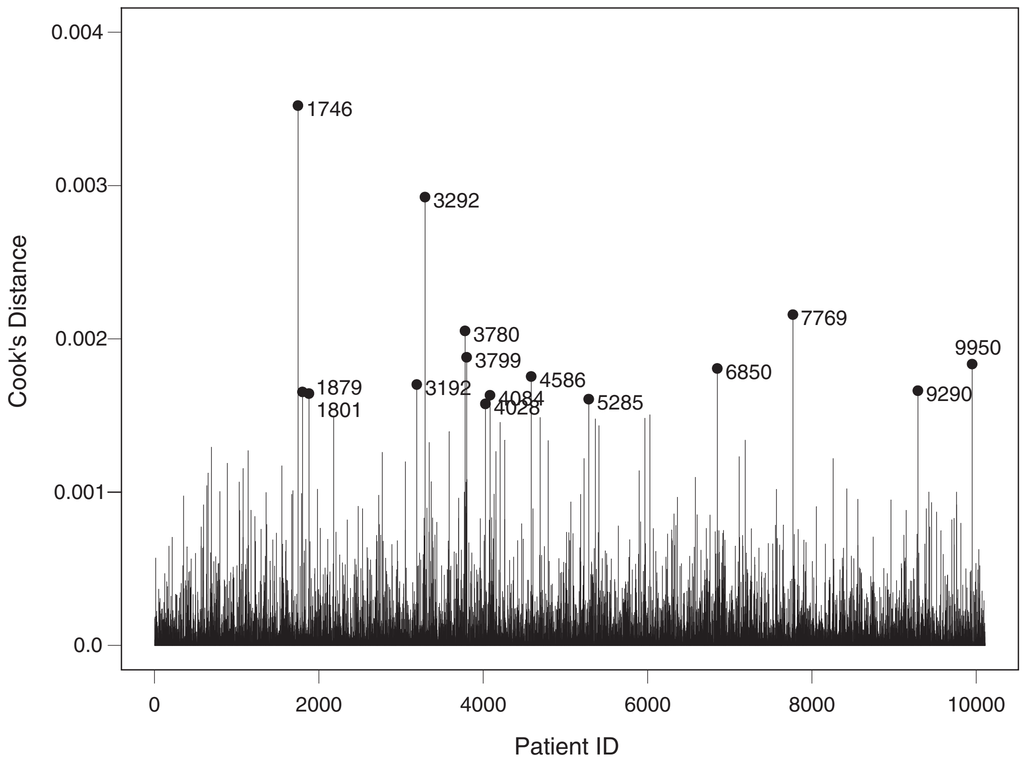

Figure 5. Plot of Cook's distance against patient ID, identifying the 15 most influential patients.

can expect that the impact of the choice of $\widehat{\mathbf{D}}$ on the influence measure will be negligible for large samples.

An important limitation of our infinitesimal influence approach and cluster deletion diagnostics is the assumption that matrix $\mathbf{V}_i$ is fixed. However, it would not be difficult to generalize the infinitesimal influence formula (6) from the score equation using the chain rule, but the influence computation would become complicated. In our influence analysis approach, we sacrifice precision for simplicity and speed. In fact, the approximations, such as (14), seem to work remarkably well in our example. For example, the error of approximation, $(\widehat{\beta}_{(5.3)} - \widehat{\beta}_{bf})/\widehat{\beta} \times 100$ per cent, where $\widehat{\beta}_{(5.3)}$ is computed by formula (14), $\widehat{\beta}_{bf}$ is computed from re-estimation, and $\widehat{\beta}$ is the value of the coefficient from Table I, is less than 0.05 per cent. Thus, the relative difference between a straightforward cluster deletion with RML re-estimation, and that derived from our approximation (14) is less than 0.05 per cent but much faster.

In this paper, we focus on influence for the fixed-effects parameters. Additional work would be required to extend these measures to the variance parameters. In that case, infinitesimal deletion diagnostics based on formula (13) where the matrix $\mathbf{V}_i$ is not fixed, would be useful because otherwise, closed-form formulae exist only when data are balanced [13, 15].

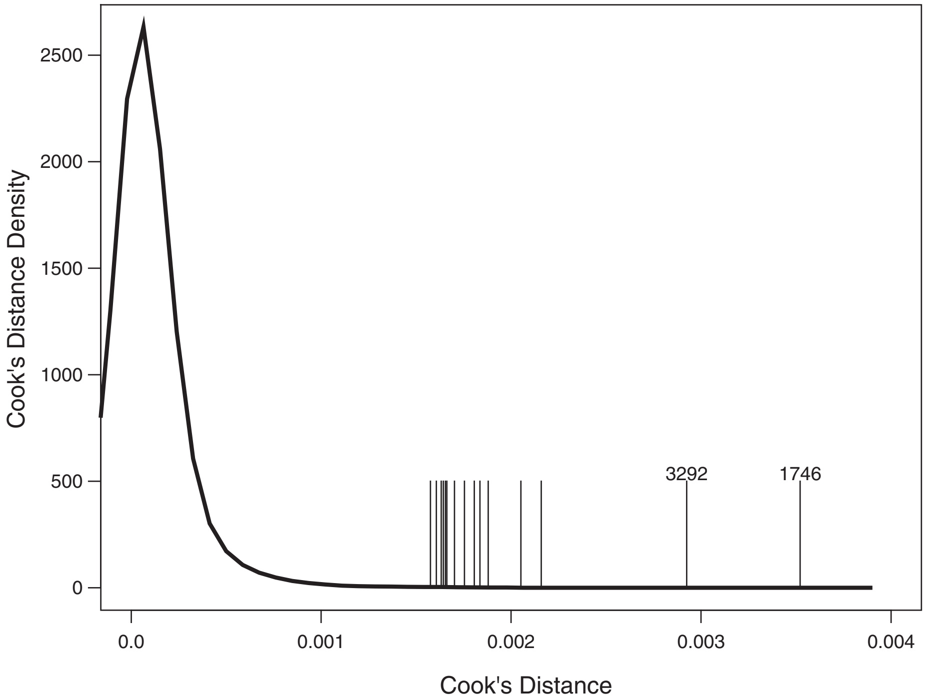

Figure 6. Density estimation for the Cook's distance. This influence measure has a long tail indicating the presence of several heavy outliers—two of them are shown at right.

With longitudinal clustered data, the proposed methods can be used to analyse the influence of clusters and/or influential observations within clusters. Different influence measures may detect different influential observations/clusters. After influential clusters or observations are detected, the undesired effects of influential observations can be attenuated by either (i) re-estimating the model after dropping influential clusters or observations, or (ii) applying robust techniques such as those based on M-estimation [15, 19, 26, 27].

Finally, all the proposed new measures have a direct interpretation in terms of the effects on a parameter of interest, and collapse to the familiar linear regression measure when there are no random effects. There are many types of diagnostic measures where one cannot rigorously prove that one measure is better than another just because they are exploratory. The value of graphical support for diagnostic analysis is always useful.

APPENDIX A

A.1. Derivation of (9)

Using previous notation (2), we denote $\widehat{\boldsymbol{\beta}}_{(i)}$ as the MLE with the ith cluster deleted, namely

$$\widehat{\boldsymbol{\beta}}_{(i)} = [\mathbf{M} - \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i]^{-1}[\mathbf{s} - \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i], \quad i = 1, \ldots, N \tag{A1}$$

Using the formula

$$(\mathbf{M} - \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1} = \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{X}_i'(\mathbf{V}_i - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i')^{-1}\mathbf{X}_i\mathbf{M}^{-1}$$

we obtain

$$\begin{aligned}
&[\mathbf{M} - \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i]^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1} \\
&\quad = [\mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{X}_i'(\mathbf{V}_i - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i')^{-1}\mathbf{X}_i\mathbf{M}^{-1}]\mathbf{X}_i'\mathbf{V}_i^{-1} \\
&\quad = \mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1} + \mathbf{M}^{-1}\mathbf{X}_i'(\mathbf{V}_i - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i')^{-1}\mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1} \\
&\quad = \mathbf{M}^{-1}\mathbf{X}_i'(\mathbf{V}_i - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i')^{-1}[(\mathbf{V}_i - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i')\mathbf{V}_i^{-1} + \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}] \\
&\quad = \mathbf{M}^{-1}\mathbf{X}_i'(\mathbf{V}_i - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i')^{-1}
\end{aligned}$$

Applying this formula to (A1) yields

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{(i)} &= \boldsymbol{\beta} + \mathbf{M}^{-1}\mathbf{X}_i'(\mathbf{V}_i - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i')^{-1}\mathbf{X}_i\boldsymbol{\beta} - \mathbf{M}^{-1}\mathbf{X}_i'(\mathbf{V}_i - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i')^{-1}\mathbf{y}_i \\
&= \boldsymbol{\beta} - \mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{I} - \mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1})^{-1}\mathbf{r}_i \\
&= \boldsymbol{\beta} - \mathbf{M}^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{I} - \mathbf{H}_{i1})^{-1}\mathbf{r}_i
\end{aligned}$$

where $\mathbf{H}_{i1}$ is the leverage matrix given by (4).

A.2. Derivation of (14)

Let the $m \times 1$ vector $\boldsymbol{\beta}$ and the $K \times 1$ vector $\boldsymbol{\delta}$ be defined as solutions to the system of non-linear equations,

$$\mathbf{F}_1(\boldsymbol{\beta}, \boldsymbol{\delta}; w) = \mathbf{0}, \quad \mathbf{F}_2(\boldsymbol{\beta}, \boldsymbol{\delta}; w) = \mathbf{0} \tag{A2}$$

where $\mathbf{F}_1$ and $\mathbf{F}_2$ are the m- and K-vector functions, respectively, and w is a scalar. The solution $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(w)$ and $\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\delta}}(w)$ are functions of w. We want to find $\mathrm{d}\widehat{\boldsymbol{\beta}}/\mathrm{d}w$, implicitly defined by (A2). Denoting

$$\mathbf{A}_{11} = \frac{\partial \mathbf{F}_1}{\partial \boldsymbol{\beta}}, \quad \mathbf{A}_{12} = \frac{\partial \mathbf{F}_1}{\partial \boldsymbol{\delta}}, \quad \mathbf{A}_{21} = \frac{\partial \mathbf{F}_2}{\partial \boldsymbol{\beta}}, \quad \mathbf{A}_{22} = \frac{\partial \mathbf{F}_2}{\partial \boldsymbol{\delta}}$$

as $m \times m$, $m \times K$, $K \times m$, $K \times K$ matrices, and differentiating (A2) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, we obtain

$$\begin{aligned}
\mathbf{A}_{11}\frac{\mathrm{d}\boldsymbol{\beta}}{\mathrm{d}w} + \mathbf{A}_{12}\frac{\mathrm{d}\boldsymbol{\delta}}{\mathrm{d}w} &= -\frac{\partial \mathbf{F}_1}{\partial w} \\
\mathbf{A}_{21}\frac{\mathrm{d}\boldsymbol{\delta}}{\mathrm{d}w} + \mathbf{A}_{22}\frac{\mathrm{d}\boldsymbol{\beta}}{\mathrm{d}w} &= -\frac{\partial \mathbf{F}_2}{\partial w}
\end{aligned} \tag{A3}$$

The solutions for $\mathrm{d}\boldsymbol{\beta}/\mathrm{d}w$ and $\mathrm{d}\boldsymbol{\delta}/\mathrm{d}w$ yield the required derivatives.

We apply this fact to the solutions of the score equations (MLE) where $\mathbf{F}_1 = \partial l/\boldsymbol{\beta}$, $\mathbf{F}_2 = \partial l/\partial\boldsymbol{\delta}$ and l is the log-likelihood function (12) with $\boldsymbol{\delta} = (\sigma^2, \text{vech}'(\mathbf{D}), \boldsymbol{\gamma})'$. In system (A3), $\mathbf{A}_{11} = -2\mathbf{M}$ and $\mathbf{A}_{12} = \mathbf{A}'_{21} = \partial^2 l/\partial\boldsymbol{\beta}\partial\boldsymbol{\delta}$, and $\partial\mathbf{F}_1/\partial w = \partial l_i/\partial\boldsymbol{\beta}$.

But it is known that $E(\partial^2 l/\partial\boldsymbol{\beta}\partial\boldsymbol{\delta}) = \mathbf{0}$ [28] so that $\mathbf{A}_{12}$ is close to zero in large samples, and we arrive at the approximation (14).

A.3. Derivation of (15)

Up to a constant, the log-likelihood function for model (1) takes the form

$$l(\boldsymbol{\beta}, \sigma^2, \mathbf{D}) = -\frac{1}{2}\sum_{i=1}^{N}\{n_i \ln \sigma^2 + [\ln|\mathbf{V}_i| + \sigma^{-2}\mathbf{r}'_i\mathbf{V}_i^{-1}\mathbf{r}_i]\} \tag{A4}$$

where $\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}$ is the $n_i \times 1$ residual vector and $\mathbf{V}_i = \mathbf{I} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}'_i$ is the $n_i \times n_i$ scaled covariance matrix of $\mathbf{y}_i$. The $(1,1)$th block of the inverse information matrix $\mathscr{I}^{-1}$ is the covariance matrix of $\widehat{\boldsymbol{\beta}}$, namely $\text{cov}(\boldsymbol{\beta}) = \sigma^2\mathbf{M}^{-1}$ [29, 18, 30]. To a first-order approximation, the log-likelihood function changes with $\boldsymbol{\beta}$ as

$$\boldsymbol{\Delta}_{i,\boldsymbol{\beta}} = \frac{\partial l_i}{\partial\boldsymbol{\beta}} = \sigma^{-2}\mathbf{X}'_i\mathbf{V}_i^{-1}\mathbf{r}_i$$

Hence, for the local influence approach, the effect of deletion of the ith cluster is approximated by

$$C_{i,\beta} = \boldsymbol{\Delta}'_{i,\boldsymbol{\beta}}\,\text{cov}(\boldsymbol{\beta})\boldsymbol{\Delta}_{i,\boldsymbol{\beta}} = \mathbf{r}'_i\mathbf{V}_i^{-1}\mathbf{X}_i\mathbf{M}^{-1}\mathbf{X}'_i\mathbf{V}_i\mathbf{r}_i$$

We can estimate $C_{i,\beta}$ by replacing unknown values with consistent estimates.

ACKNOWLEDGEMENTS

We thank the reviewers for relevant comments that lead to an improvement of the paper.

REFERENCES

1. Cook RD, Weisberg S. *Residuals and Influence in Regression.* Chapman & Hall: New York, 1982.
2. Chatterjee S, Hadi AS. Influential observations, high leverage points, and outliers in linear regression. *Statistical Science* 1986; **1**(3):379–416.
3. Rao CR, Toutenburg H. *Linear Models. Least Squares and Alternatives.* Springer: New York, 1997.
4. Laurent RSST, Cook RD. Leverage and superleverage in nonlinear regression. *Journal of American Statistical Association* 1992; **87**(420):985–990.
5. Laurent RSST, Cook RD. Leverage, local influence and curvature in nonlinear regression. *Biometrika* 1993; **80**(1):99–106.
6. Beckman RJ, Trussel HJ. The distribution of an arbitrary studentized residual and the effects of updating in multiple regression. *Journal of the American Statistical Association* 1974; **69**(345):199–201.
7. Miller RG. An unbalanced jackknife. *Annals of Statistics* 1974; **2**(5):880–891.
8. Cook RD. Detection of influential observation in linear regression. *Technometrics* 1977; **19**(1):5–18.
9. Pregibon D. Logistic regression diagnostic. *Annals of Statistics* 1981; **4**(4):705–724.
10. Cook RD. Assessment of local influence (with discussion). *Journal of the Royal Statistical Society*, *Series B* 1986; **48**(2):139–169.
11. Lesaffre E, Verbeke G. Local influence in linear mixed models. *Biometrics* 1998; **54**(2):570–582.
12. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2000.
13. Pan JX, Fang KT. *Growth Curve Models and Statistical Diagnostics*. Springer: New York, 2002.
14. Frans EST, Ouwens MJN, Berger MPF. Detection of influential observations in longitudinal mixed effects regression models. *The Statistician* 2001; **50**(3):271–284.

15. Demidenko E. *Mixed Models*: *Theory and Applications*. Wiley: Hoboken, U.S.A., 2004, in press.
16. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**(4):963–974.
17. McCulloch CE, Searle SR. *Generalized*, *Linear and Mixed Models*. Wiley: New York, 2001.
18. Vonesh EF, Chinchilli VM. *Linear and Nonlinear Models for the Analysis of Repeated Measurements.* Marcel Dekker: New York, 1977.
19. Huber PJ. *Robust Statistics*. Wiley: New York, 1967.
20. Christensen R, Pearson LM, Johnson W. Case-deletion diagnostics for mixed models. *Technometrics* 1992; **34**(1):38–45.
21. Banerjee M, Frees EW. Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association* 1997; **92**(439):999–1005.
22. Demidenko E, Stukel TA. Efficient estimation of general linear mixed effects models. *Journal of Statistical Planning and Inference* 2002; **104**(1):197–219.
23. Lindstrom MJ, Bates DM. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* 1988; **83**(404):1014–1022.
24. Beckman RJ, Nachtsheim CJ, Cook RD. Diagnostics for mixed-model analysis of variance. *Technometrics* 1987; **29**(4):413–425.
25. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 1: the content, quality and accessibility of care. *Annals of Internal Medicine* 2003; **138**(4):273–287.
26. Zaslavsky AM, Schenker N, Berlin TR. Downweighting influential clusters in surveys: application to the 1990 post enumeration survey. *Journal of American Statistical Association* 2001; **96**(455):858–869.
27. Yeap BY, Davidian M. Robust two-stage estimation in hierarchical nonlinear models. *Biometrics* 2001; **57**(1):266–272.
28. Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 1986; **42**(3):805–820.
29. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurement Data.* London: Chapman & Hall, 1995.
30. Pinheiro JC, Bates DM. *Mixed Effects Models in S-Plus.* New York: Springer, 2000.

