



Published in final edited form as:

J Comput Graph Stat. 2012 ; 21(1): 253–271. doi:10.1198/jcgs.2011.10139.

Bayesian Case Influence Measures for Statistical Models with Missing Data

Hongtu Zhu,

Department of Biostatistics, University of North Carolina at Chapel Hill

Joseph G. Ibrahim,

Department of Biostatistics, University of North Carolina at Chapel Hill

Hyunsoon Cho*, and

Department of Biostatistics, University of North Carolina at Chapel Hill

Niansheng Tang

Department of Statistics, Yunnan University

Abstract

We examine three Bayesian case influence measures including the ϕ -divergence, Cook's posterior mode distance and Cook's posterior mean distance for identifying a set of influential observations for a variety of statistical models with missing data including models for longitudinal data and latent variable models in the absence/presence of missing data. Since it can be computationally prohibitive to compute these Bayesian case influence measures in models with missing data, we derive simple first-order approximations to the three Bayesian case influence measures by using the Laplace approximation formula and examine the applications of these approximations to the identification of influential sets. All of the computations for the first-order approximations can be easily done using Markov chain Monte Carlo samples from the posterior distribution based on the full data. Simulated data and an AIDS dataset are analyzed to illustrate the methodology.

Keywords

Case influence measures; Cook distance; First-order approximation; ϕ -divergence; Markov chain Monte Carlo

1 Introduction

One of the main goals of any statistical analysis is to assess model assumptions and model fit. Towards this goal, it is critical to assess the influence of individual cases (or generally, a set of observations) on an analysis and to identify influential observations (or sets of observations) and/or outliers (Cook, 1977; Cook and Weisberg, 1982; McCulloch, 1989; Geisser, 1975, 1993). In Bayesian analysis, considerable research has been devoted to

*We thank the editor, an associate editor and a referee for many valuable suggestions. This work was supported in part by NSF SES-06-43663 and BCS-08-26844 and NIH grants UL1-RR025747-01, P01CA142538-01, AG033387, MH086633, GM70335, and CA74015 to Drs. Zhu and Ibrahim, and NSFC grants 10561008, 1076104, and 10961026 and NCET grant NCET-07-0737 to Dr. Tang.

Supplemental materials for the article are available online.

C++ Code The supplemental files for this article include C++ programs which can be used to replicate the simulation study and real data analysis included in the article. Please read file README contained in the zip file for more details. (ZhuJoeTang.zip, zip archive)
Appendix The supplemental files include the Appendix which all variables in the AIDS dataset and the proofs of Theorems 1 and 2. (cho2-jcgsApp.pdf)

developing single case influence measures for various specific statistical models including generalized linear models, time series models, and survival models (Johnson and Geisser, 1983; Johnson, 1985; Pettit, 1986; Kass et al., 1989; Carlin and Polson, 1991; Gelfand et al., 1992; Geisser, 1993; Blyth, 1994; Peng and Dey, 1995; Christensen, 1997; Bradlow and Zaslavsky, 1997). The influence of an individual observation or set of observations is often assessed by deleting the observation (or set of observations) and then comparing the posterior (or predictive) distribution based on the full data compared to that of posterior (or predictive) distribution based on the deleted observation (observations).

There are four major types of Bayesian case influence measures. These are posterior probabilities of outlying sets, posterior outlier statistics, predictive diagnostics, and posterior diagnostics for identifying outliers and influential points (Zhu et al., 2010). Computing posterior probabilities of outlying sets is conceptually simple (Box and Tiao, 1968), but it is difficult to implement computationally in most models with missing data. So far, this approach is limited to several simple regression models, such as the classical linear regression model with conjugate priors (Abraham and Box, 1979). The posterior outlier statistic is based on using the posterior distribution of an outlier statistic, such as the raw residual, to define outliers and calculate the posterior probability that an observation is an outlier. This method is computationally simple and has been further extended to generalized linear models, survival models, latent variable models, state space models, and many others (Chaloner, 1991; Albert and Chib, 1993; Lee, 2007). Predictive diagnostics assess the discordance of a set of observations based on their predictive distribution (Gelfand et al., 1992; Geisser, 1993; Gelfand and Dey, 1994). Predictive diagnostics are also conceptually simple, but computing them can be difficult for regression models with missing data. In contrast to predictive diagnostics, posterior diagnostics compare the posterior distributions of the parameters given the complete data and the reduced data (Csiszár, 1967; Weiss and Cook, 1992).

Despite the extensive literature on Bayesian diagnostics for various types of models, very little has been done on systematically examining Bayesian case influence measures using case deletion, namely the ϕ -divergence, *Cook's posterior mode distance* and *Cook's posterior mean distance*, in statistical models for both dependent and independent data in the presence of missing data. We refer the reader to a review of Bayesian diagnostics in Zhu et al. (2010) and Peng and Dey (1995). Computationally, as shown later, it can be very difficult to directly compute such Bayesian case influence measures for many complex models with missing data (Molenberghs and Kenward, 2007; Molenberghs and Verbeke, 2005; Lee, 2007; Ibrahim et al., 2005; Daniels and Hogan, 2008; Little, 1992; Little and Rubin, 2002; Ibrahim and Molenberghs, 2009; Skrondal and Rabe-Hesketh, 2004). For example, in the real data analysis presented in Section 3.2, we present a Bayesian diagnostic analysis for a complex Bayesian structural equations model with nonignorable missing data, for which it is infeasible to compute the exact values of these Bayesian case influence measures. This setting thus facilitates the need for deriving computationally feasible approximations for these Bayesian case influence measures.

The aims of this paper are to systematically examine the above-mentioned Bayesian case influence measures based on case deletion, to derive their first-order approximations, and to evaluate their roles in detecting a set of influential observations for a variety of regression models with missing data. By using a Laplace approximation (Kass et al., 1990; Tierney et al., 1989), we show that under some mild conditions, the first-order approximations hold for a large class of statistical models for both dependent and independent data within the Bayesian framework. We extensively examine the accuracy of these first-order approximations for the three Bayesian case influence measures using both theoretical results and simulation studies. Specifically, we show that the first-order approximations are quite

accurate and all of the computations for the first-order approximations can be easily numerically computed using Markov chain Monte Carlo (MCMC) samples from the posterior distribution based on the full data.

The rest of this paper is organized as follows. In Section 2, we introduce the three Bayesian case influence measures and propose computational formulas for them. We derive two first-order approximations to the Bayesian case influence measures. In Section 3, we illustrate the proposed methodology for latent variable models with missing data. We conclude the paper with some discussions in Section 4.

2 Methods

2.1 Bayesian Case Influence Measures

Let $p(\mathbf{Y}|\boldsymbol{\theta})$ be the probability function for a random vector $\mathbf{Y}^T = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_n^T)$, parameterized by an unknown parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ in an open subset Θ of R^p . Moreover, the dimension of $\mathbf{Y}_i = (y_{i1}, \dots, y_{im_i})^T$, denoted by m_i , can vary across all i . For example, in longitudinal studies and mixed models, m_i is the number of observations in each cluster and this may vary significantly across the clusters. Let $p(\boldsymbol{\theta})$ be the prior distribution of $\boldsymbol{\theta}$. The posterior distribution for the full data \mathbf{Y} is given by $p(\boldsymbol{\theta}|\mathbf{Y}) \propto p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$.

We are interested in assessing the influence of deleting a set of observations, denoted by S , on posterior inferences regarding $\boldsymbol{\theta}$. Let $N = \sum_{i=1}^n m_i$ and N_S be, respectively, the total number of observations and the number of observations in the set S . A subscript ‘[S]’ denotes the relevant quantity with all observations in S deleted. For instance, if $S = \{i\}$, then $\mathbf{Y}_{[S]}$ is the corresponding observed data with all of \mathbf{Y}_i deleted, whereas if $S = \{i_1, i_2\}$, then $\mathbf{Y}_{[S]}$ is the corresponding observed data with \mathbf{Y}_{i_1} and \mathbf{Y}_{i_2} deleted. Furthermore, we may set $S = \{i_1, \dots, i_k\}$ and $S = \{(i_1, j_1), \dots, (i_k, j_k)\}$ to allow for more complicated case deletion schemes. Let \mathbf{Y}_S denote a subsample of \mathbf{Y} consisting of all the observations in S and let $\mathbf{Y}_{[S]}$ denote a subsample of \mathbf{Y} with all observations in \mathbf{Y}_S deleted. The posterior distribution for a subsample of the data \mathbf{Y} is given by $p(\boldsymbol{\theta}|\mathbf{Y}_{[S]}) \propto p(\mathbf{Y}_{[S]}|\boldsymbol{\theta})p(\boldsymbol{\theta})$.

Now, we examine three types of Bayesian case influence measures based on case deletion. The first type is the ϕ -influence of $\mathbf{Y}_{[S]}$, defined by

$$D_\phi(S) = \int \phi(R_{[S]}(\boldsymbol{\theta}))p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta}, \quad (1)$$

where $R_{[S]}(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{Y}_{[S]})/p(\boldsymbol{\theta}|\mathbf{Y})$ and $\phi(\cdot)$ is a convex function with $\phi(1) = 0$ (Weiss and Cook, 1992; Weiss, 1996). $D_\phi(S)$ directly measures the distance (discrepancy) between two posterior distributions $p(\boldsymbol{\theta}|\mathbf{Y}_{[S]})$ and $p(\boldsymbol{\theta}|\mathbf{Y})$ (Csiszár, 1967; Weiss and Cook, 1992) and a large value of $D_\phi(S)$ corresponds to a set of influential observations. Various forms of $\phi(\cdot)$ have been widely considered in the literature (Kass et al., 1989; Weiss and Cook, 1992; Blyth, 1994; Peng and Dey, 1995; Weiss, 1996). For instance, $\phi(\cdot)$ can be chosen to be $\phi_\alpha(u)$, which is defined by $4\{1 - u^{(1+\alpha)/2}\}/\{1 - \alpha^2\}$ for $\alpha \neq \pm 1$, $u \log(u)$ for $\alpha = 1$, and $-\log(u)$ for $\alpha = -1$. In particular, $\phi_1(\cdot)$ and $\phi_{-1}(\cdot)$ lead to the Kullback-Leibler divergence (K-L divergence); moreover, $\phi(u) = \phi_1(u) + \phi_{-1}(u)$ leads to the symmetric K-L divergence. The L_1 -distance and the χ^2 -divergence correspond to $\phi(u) = 0.5|u - 1|$ and $\phi(u) = (u - 1)^2$, respectively (Weiss, 1996).

The second Bayesian influence measure assesses the discrepancy between the posterior mode of $\boldsymbol{\theta}$ with and without the i th case (Cook and Weisberg, 1982). We call this measure *Cook's posterior mode distance*. Specifically, we define the posterior modes of $\boldsymbol{\theta}$ for the full

sample \mathbf{Y} and a subsample $\mathbf{Y}_{[S]}$ as $\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}_{[S]} = \operatorname{argmax}_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\mathbf{Y}_{[S]})$, respectively. Then, Cook's posterior mode distance for comparing \mathbf{Y} and $\mathbf{Y}_{[S]}$, denoted by $\text{CP}(S)$, can be defined as follows:

$$\text{CP}(S) = (\hat{\boldsymbol{\theta}}_{[S]} - \hat{\boldsymbol{\theta}})^T \mathbf{G}_{\boldsymbol{\theta}} (\hat{\boldsymbol{\theta}}_{[S]} - \hat{\boldsymbol{\theta}}), \quad (2)$$

where $\mathbf{G}_{\boldsymbol{\theta}}$ is chosen to be $-\partial_{\boldsymbol{\theta}}^2 \log p(\boldsymbol{\theta}|\mathbf{Y}) = -\partial_{\boldsymbol{\theta}}^2 \log p(\mathbf{Y}|\boldsymbol{\theta}) - \partial_{\boldsymbol{\theta}}^2 \log p(\boldsymbol{\theta})$ evaluated at $\hat{\boldsymbol{\theta}}$, where $\partial_{\boldsymbol{\theta}}^2$ represents the second-order derivative with respect to $\boldsymbol{\theta}$. If we consider a uniform improper prior for $\boldsymbol{\theta}$, then $\text{CP}(S)$ reduces to the well-known Cook's distance for deleting a set of observations (Cook and Weisberg, 1982). A large value of $\text{CP}(S)$ implies more influence of the set S on the posterior mode.

The third type of Bayesian influence measure assesses the distance between the posterior mean of $\boldsymbol{\theta}$ with and without the observations in S . We define the posterior mean of $\boldsymbol{\theta}$ for the full sample \mathbf{Y} and a subsample $\mathbf{Y}_{[S]}$ as $\tilde{\boldsymbol{\theta}} = \int \boldsymbol{\theta} \cdot p(\boldsymbol{\theta}|\mathbf{Y}) d\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}_{[S]} = \int \boldsymbol{\theta} \cdot p(\boldsymbol{\theta}|\mathbf{Y}_{[S]}) d\boldsymbol{\theta}$, respectively. Cook's posterior mean distance for deleting the observations in the set S , denoted by $\text{CM}(S)$, can then be defined as follows:

$$\text{CM}(S) = (\tilde{\boldsymbol{\theta}}_{[S]} - \tilde{\boldsymbol{\theta}})^T \mathbf{W}_{\boldsymbol{\theta}} (\tilde{\boldsymbol{\theta}}_{[S]} - \tilde{\boldsymbol{\theta}}), \quad (3)$$

where $\mathbf{W}_{\boldsymbol{\theta}}$ is chosen to be the inverse of the full-data posterior covariance matrix of $\boldsymbol{\theta}$. A large value of $\text{CM}(S)$ corresponds to an influential set S regarding the posterior mean.

Although all three Bayesian case influence measures assess the influence of a set of observations, there is a conceptual difference among those measures. $D_{\phi}(S)$ quantifies the effects of deleting a set of observations on the overall posterior distribution, whereas $\text{CP}(S)$ and $\text{CM}(S)$ quantify the effects of deleting a set of observations on the posterior mode and the posterior mean of $\boldsymbol{\theta}$, respectively. Since $D_{\phi}(S)$ measures the overall difference between $p(\boldsymbol{\theta}|\mathbf{Y})$ and $p(\boldsymbol{\theta}|\mathbf{Y}_{[S]})$, and such a difference may include shape, mode, mean etc., $D_{\phi}(S)$ can be more sensitive to some changes of the posterior distributions other than the posterior mean or posterior mode due to the deletion of the observations in S compared with $\text{CP}(S)$ and $\text{CM}(S)$. However, compared with $D_{\phi}(S)$, $\text{CP}(S)$ and $\text{CM}(S)$ may be more sensitive to a change in the posterior mean or posterior mode.

2.2 Computational Formula and its Difficulties

When $p(\mathbf{Y}|\boldsymbol{\theta})$ is relatively easy to compute, all three Bayesian case influence measures can be computed using only MCMC samples from the full posterior distribution, $p(\boldsymbol{\theta}|\mathbf{Y})$. We define $p_S(\boldsymbol{\theta})$, the ratio of likelihoods with and without the observations in S as

$$p_S(\boldsymbol{\theta}) = \frac{p(\mathbf{Y}|\boldsymbol{\theta})}{p(\mathbf{Y}_{[S]}|\boldsymbol{\theta})} = p(\mathbf{Y}_S|\mathbf{Y}_{[S]}, \boldsymbol{\theta}), \quad (4)$$

which is the conditional distribution of \mathbf{Y}_S given $\mathbf{Y}_{[S]}$. Then, we have

$$p(\boldsymbol{\theta}|\mathbf{Y}_{[S]}) = [p_S(\boldsymbol{\theta})]^{-1} p(\mathbf{Y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) / \int [p_S(\boldsymbol{\theta})]^{-1} p(\mathbf{Y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Thus, following Weiss (1996), the computational formula for $D_{\phi}(S)$ can be obtained as

$$D_{\phi}(S) = E_{\theta|Y} \left[\phi \left(\frac{[p_S(\theta)]^{-1}}{E_{\theta|Y} \{ [p_S(\theta)]^{-1} \}} \right) \right], \quad (5)$$

where $E_{\theta|Y}$ denotes the expectation taken with respect to the posterior distribution $p(\theta|Y)$. Specifically, for the K-L divergence ($\phi(u) = -\log(u)$), the computational formula is given by $D_{\phi}(S) = \log E_{\theta|Y} \{ [p_S(\theta)]^{-1} \} + E_{\theta|Y} \{ \log [p_S(\theta)] \}$. It is well recognized that the accuracy of approximating $D_{\phi}(S)$ depends heavily on the variability of $p_S(\theta)$ (Epifani et al., 2008; Peruggia, 1997).

To compute $CP(S)$, we need to evaluate $\hat{\theta}$ and $\hat{\theta}_S$. In general, the posterior mode of θ does not have a closed analytic form, thus we have to rely on iterative methods such as Newton-Raphson to obtain $\hat{\theta}$ and $\hat{\theta}_{[S]}$. However, this can be computationally intensive for most models, such as state space models. G_{θ} in $CP(S)$ can be analytically obtained by evaluating $J_N(\theta) = -\partial_{\theta}^2 \log p(\theta|Y) = -\partial_{\theta}^2 \log p(Y|\theta) - \partial_{\theta}^2 \log p(\theta)$ at $\hat{\theta}$.

Since we can write $\tilde{\theta} = E_{\theta|Y}(\theta)$ and

$$\tilde{\theta}_{[S]} = E_{\theta|Y} \{ \theta \cdot [p_S(\theta)]^{-1} \} / E_{\theta|Y} \{ [p_S(\theta)]^{-1} \}, \quad (6)$$

we can easily compute $CM(S)$ using MCMC samples from the full posterior distribution, $p(\theta|Y)$. Specifically, the posterior mean of θ , denoted $\tilde{\theta}$, can be obtained directly by averaging the MCMC samples and W_{θ} can be analytically obtained by evaluating $J_N(\theta)$ at $\tilde{\theta}$. Furthermore, G_{θ} can be approximated by the inverse of the posterior covariance matrix, obtained from the MCMC samples. Based on the above discussion, computing (5) and (6) strongly depends on the the computation of $p_S(\theta) = p(Y_S|Y_{[S]}, \theta)$.

It can be computationally quite cumbersome to approximate $p_S(\theta)$ in the presence of missing data, which therefore makes the computation of the three Bayesian case influence measures infeasible. To see this fact, denote the missing data by $Y_{mis} = (Y_{1,mis}^T, \dots, Y_{n,mis}^T)^T$ and the complete data by $Y_{com} = (Y_{1,com}^T, \dots, Y_{n,com}^T)^T$, in which $Y_{i,com} = (Y_{i,mis}, Y_i)$ for $i = 1, \dots, n$. Let $p(Y_{com}|\theta)$ be the probability function for Y_{com} . We define $Y_{com,[S]} = (Y_{[S]}, Y_{mis})$ as the complete data after deleting all observations in Y_S and $p(Y_{com,[S]}|\theta)$ is the probability function for $Y_{com,[S]}$ such that $\int p(Y_{com,[S]}|\theta) dY_{mis} = p(Y_{[S]}|\theta)$. This kind of model structure is very general and subsumes most commonly used models, such as GLMs with missing responses and/or covariates and random-effects models (Ibrahim et al., 2005, 2010; Zhu et al., 2001; Molenberghs and Kenward, 2007; Molenberghs and Verbeke, 2005; Lee, 2007; Ibrahim et al., 2005; Daniels and Hogan, 2008; Little, 1992; Little and Rubin, 2002; Ibrahim and Molenberghs, 2009; Skrondal and Rabe-Hesketh, 2004). With missing data, the primary computational challenge lies in the computation of $p_S(\theta)$, because

$$p_S(\theta) = \frac{\int p(Y, Y_{mis}|\theta) dY_{mis}}{\int p(Y_{[S]}, Y_{mis}|\theta) dY_{mis}} \quad (7)$$

typically involves high-dimensional integrals.

Example 1—To illustrate the methodological development, we consider n independent observations $\{Y_{i,com} = (\mathbf{x}_i, \mathbf{z}_i, r_i, y_i), i = 1, \dots, N\}$, where y_i is the response variable, \mathbf{x}_i is a $p_1 \times 1$ vector of completely observed covariates, and \mathbf{z}_i is a $p_2 \times 1$ vector of partially observed

covariates. Moreover, let $z_{mis,i}$ and $z_{obs,i}$ denote the missing and observed components of z_j , respectively. Let r_j be a $p_2 \times 1$ vector, whose j^{th} component, r_{ij} , equals 1 if z_{ij} is observed, and 0 if z_{ij} is missing. We assume that $p(\mathbf{x}_j, z_j, r_j, y_j | \theta) = p(y_j | \mathbf{x}_j, z_j, \theta) p(\mathbf{x}_j, z_j | \theta) p(r_j | y_j, \mathbf{x}_j, z_j, \theta)$, where θ denotes the vector of unknown parameters.

We assume a generalized linear model for $p(y_j | \mathbf{x}_j, z_j, \beta, \tau)$ given by

$$p(y_i | x_i, z_i, \beta, \tau) = \exp \left\{ a_i^{-1}(\tau) [y_i \eta_i(\beta) - b(\eta_i(\beta))] + c(y_i, \tau) \right\} \quad (8)$$

for $i = 1, \dots, n$, where $a_i(\cdot)$, $b(\cdot)$ and $c(\cdot, \cdot)$ are known functions, $\eta_i = \eta(\mu_i)$ and $\mu_i = g((x_i', z_i')\beta)$, in which $g(\cdot)$ is a known link function, $\beta = (\beta_1, \dots, \beta_p)'$ and $p = p_1 + p_2$. We assume that

$$p(x_i, z_i | \alpha) = p(z_{ip_2} | z_{i,p_2-1}, \dots, z_{i1}, x_i, \alpha) \times \dots \times p(z_{i1} | x_i, \alpha) p(x_i | \alpha). \quad (9)$$

Similarly, we model the missing-data mechanism $p(r_j | y_j, \mathbf{x}_j, z_j, \xi)$ as

$$p(r_i | y_i, x_i, z_i, \xi) = p(r_{ip_2} | r_{i,p_2-1}, \dots, r_{i1}, y_i, x_i, z_i, \xi) \times \dots \times p(r_{i1} | y_i, x_i, z_i, \xi). \quad (10)$$

Here, $\theta = (\beta, \tau, \alpha, \xi)$. To carry out a Bayesian analysis, we need to specify a prior for θ . Following Huang, Chen and Ibrahim (2005), we specify a prior of θ such that $p(\theta) = p(\tau)p(\beta)p(\xi)p(\alpha)$.

Now we consider the deletion of the i -th observation $(\mathbf{x}_j, z_j, r_j, y_j)$, that is $S = \{i\}$. With some calculations, we get

$$p_{\{i\}}(\theta) = \int p(y_i | x_i, z_i, \beta, \tau) p(x_i, z_i | \alpha) p(r_i | y_i, x_i, z_i, \xi) dz_{mis,i}, \quad (11)$$

which may involve intractable integrals when the dimension of $z_{mis,i}$ is relatively large. Although one may be able to use some numerical methods (e.g., the trapezoidal rule) to approximate $p_S(\theta)$, the accuracy of such approximations can be impossible to assess when the integrals are high dimensional. This setting thus requires the derivation of computationally simple approximations to these Bayesian case influence measures.

2.3 First-order Approximations

For diagnostic purposes, it is desirable to derive computationally feasible approximations to these case influence measures. We obtain the following theorems, whose detailed proofs can be found in the the supplementary document.

Theorem 1—If Assumptions C1-C5 in the supplementary document hold and N_S is bounded by a fixed constant, then we have the following results:

- a. $D_\phi(S)$ can be approximated by

$$D_\phi(S) = 0.5 \dot{\phi}(1) [\partial_\theta \log p_S(\hat{\theta})]^T [J_N(\hat{\theta})]^{-1} [\partial_\theta \log p_S(\hat{\theta})] [1 + O_p(N^{-1})], \quad (12)$$

where $\dot{\phi}(1) = \partial_u^2 \phi(u)|_{u=1}$.

- b. The one-step approximation for $\hat{\theta}_{[S]}$ is given by

$$\widehat{\theta}_{[S]} = \widehat{\theta} + O_p(N^{-1}) = \widehat{\theta} - [J_N(\widehat{\theta})]^{-1} \partial_{\theta} \log p_S(\widehat{\theta}) [1 + O_p(N^{-1})]. \quad (13)$$

c. The one-step approximation for $\widehat{\theta}_{[S]}$ is given by

$$\tilde{\theta}_{[S]} = \tilde{\theta} - [J_N(\tilde{\theta})]^{-1} \partial_{\theta} \log p_S(\tilde{\theta}) [1 + O_p(N^{-1})]. \quad (14)$$

d. $2D_{\phi}(S)/\check{\phi}(1)$, $CP(S)$, and $CM(S)$ are asymptotically equivalent, that is,

$$D_{\phi}(S) = 0.5\check{\phi}(1) \times CP(S) + O_p(N^{-2}) = 0.5\check{\phi}(1) \times CM(S) + O_p(N^{-2}). \quad (15)$$

Theorem 1 has several important implications. Theorem 1 (a) provides a theoretical and computational approximation of $D_{\phi}(S)$ as a quadratic form in $\partial_{\theta} \log p_S(\tilde{\theta})$. Theorem 1 (b) and (c) provide the one-step approximations of $\widehat{\theta}_{[S]}$ and $\tilde{\theta}_{[S]}$, which reduce the burden of computing $\widehat{\theta}_{[S]}$ and $\tilde{\theta}_{[S]}$ for each S . Moreover, to the best of our knowledge, Theorem 1 (d) is the first result that establishes a direct connection between $D_{\phi}(S)$, $CP(S)$ and $CM(S)$ for any $\phi(\cdot)$ within the Bayesian framework. In particular, for $\phi_{\alpha}(u) = -\log(u)$, it can be shown that $\partial_u^2 \phi_{\alpha}(u)|_{u=1} = 1$, which leads to $D_{\phi_{\alpha}}(S) = 0.5CP(S) + O_p(N^{-2}) = 0.5CM(S) + O_p(N^{-2})$ for all α . Furthermore, for the χ^2 -divergence and the symmetric K-L divergence, we have $\partial_u^2 \phi(u)|_{u=1} = 2$, which gives $D_{\phi}(S) = CP(S) + O_p(N^{-2}) = CM(S) + O_p(N^{-2})$. However, for $\phi(u) = 0.5|u - 1|$, because $\check{\phi}(1) = 0$ and $\phi(u)$ is not differentiable at $u = 1$, the conditions of Theorem 1 are not valid. Thus, the approximation given in Theorem 1 (a) and the equivalence among the three diagnostic measures do not hold for the L_1 -distance.

Practically, these approximations lead to computationally efficient formulas for approximating these Bayesian case influence measures. Because the first-order approximations hold for a large class of statistical models for both dependent and independent data, and in the presence of missing data within the Bayesian framework, they are reminiscent of the first-order approximations for various specific models within the frequentist framework (Christensen et al., 1992; Cook and Weisberg, 1982; Wei, 1998). For influential points, even though the accuracy of the first-order approximation may be relatively low, the first-order approximated measure can easily pick out these influential points. Thus, for diagnostic purposes, the first-order approximation may be more effective at identifying influential clusters compared with the three Bayesian case influence measures. We conduct simulation studies to investigate the performance of these first-order approximations relative to the exact formula in Section 3. See numerical comparisons in Table 4.

According to Theorem 1, to approximate these case influence measures, we only need to compute the posterior mean $\tilde{\theta}$, the observed-data information matrix $J_N(\tilde{\theta})$, $\partial_{\theta} \log p_S(\tilde{\theta})$ evaluated at $\tilde{\theta}$, and

$$AP_1(S; \tilde{\theta}) = [\partial_{\theta} \log p_S(\tilde{\theta})]^T [J_N(\tilde{\theta})]^{-1} [\partial_{\theta} \log p_S(\tilde{\theta})]. \quad (16)$$

In particular, $\tilde{\theta}$ and $J_N(\tilde{\theta})$ can be easily computed from the MCMC samples based on the full data. Specifically, $J_N(\tilde{\theta})$ can be approximated by using the Louis' formula (Louis, 1982). For most statistical models, the computation of $\partial_{\theta} \log p_S(\tilde{\theta}) = \partial_{\theta} \log p(Y|\tilde{\theta}) - \partial_{\theta} \log p(Y_{[S]}|\tilde{\theta})$ is relatively straightforward. Specifically, we have

$$\partial_{\theta} \log p(\mathbf{Y}|\tilde{\theta}) \approx \int \partial_{\theta} \log p(\mathbf{Y}_{com}|\tilde{\theta}) p(\mathbf{Y}_{mis}|\mathbf{Y}, \tilde{\theta}) d\mathbf{Y}_{mis}, \quad (17)$$

$$\partial_{\theta} \log p(\mathbf{Y}_{[S]}|\tilde{\theta}) \approx \int \partial_{\theta} \log p(\mathbf{Y}_{com,[S]}|\tilde{\theta}) p(\mathbf{Y}_{mis}|\mathbf{Y}, \tilde{\theta}) d\mathbf{Y}_{mis}. \quad (18)$$

Here, we use the fact that the posterior mean and posterior mode are asymptotically equivalent under suitable regularity conditions which are satisfied for the models considered here.

Example 1 (continued): Let's consider deletion of the i -th observation $(\mathbf{x}_i, \mathbf{z}_i, \mathbf{r}_i, y_i)$. It can be shown that

$$\partial_{\theta} \log p_{(i)}(\tilde{\theta}) = \int \partial_{\theta} [\log p(y_i|x_i, \mathbf{z}_i, \beta, \tau) + \log p(x_i, \mathbf{z}_i|\alpha) + \log p(\mathbf{r}_i|y_i, x_i, \mathbf{z}_i, \xi)]|_{\theta=\tilde{\theta}} p(\mathbf{z}_{mis,i}|y_i, x_i, \mathbf{z}_{obs,i}, \tilde{\theta}) d\mathbf{z}_{mis,i}.$$

Note that in most models with missing data, it is relatively easy to compute the first-order derivative of the complete-data log-likelihood function. Moreover, compared to the computation of the observed-data log-likelihood function, it is numerically much more stable to calculate the first-order derivative of the complete-data log-likelihood function. This is the key advantage of using (17) and (18) to approximate $AP_1(\{i\}; \tilde{\theta})$.

We now present the four key steps in computing $AP_1(S; \tilde{\theta})$ in (16).

Step 1: Using the full data \mathbf{Y} , we obtain the MCMC sample $\theta^{(j)}$ for $j = 1, \dots, J$ from $p(\mathbf{Y}|\theta)$ and estimate $\tilde{\theta} = J^{-1} \sum_{j=1}^J \theta^{(j)}$.

Step 2: We use MCMC methods to draw samples $\{\mathbf{Y}_{mis}^{(j)}: j=1, \dots, J\}$ from $p(\mathbf{Y}_{mis}|\mathbf{Y}, \tilde{\theta})$ given $\tilde{\theta}$ and \mathbf{Y} .

Step 3: For each set S , we approximate $J_N(\tilde{\theta})$ using Louis' formula and approximate $\theta \log p(\mathbf{Y}|\tilde{\theta})$ and $\theta \log p(\mathbf{Y}_{[S]}|\tilde{\theta})$ by

$$\partial_{\theta} \log p(\mathbf{Y}|\tilde{\theta}) \approx J^{-1} \sum_{j=1}^J \partial_{\theta} \log p(\mathbf{Y}_{mis}^{(j)}, \mathbf{Y}_{obs}|\tilde{\theta}), \quad (19)$$

$$\partial_{\theta} \log p(\mathbf{Y}_{[S]}|\tilde{\theta}) \approx J^{-1} \sum_{j=1}^J \partial_{\theta} \log p(\mathbf{Y}_{mis}^{(j)}, \mathbf{Y}_{[S]}|\tilde{\theta}). \quad (20)$$

Step 4: Approximate $AP_1(S; \tilde{\theta})$ using equation (16) for each set S .

Although we have systematically examined the deletion of a relatively small number of observations, it is common to delete relatively large numbers of observations for clustered data. Specifically, unbalanced clustered data are commonly collected from familial and longitudinal studies and we may be interested in deleting all the observations in a cluster, whose number may be comparable with the total number of observations N (Wang et al., 1999). We now obtain the following theorem for large cluster sizes.

Theorem 2—If Assumptions C1, C2, C3, C4' and C5 in the supplementary document hold and $N_S \rightarrow \infty$ and $N_S/N \rightarrow \gamma \in [0, 1)$, then we have the following results:

- a. The one-step approximation for $\hat{\theta}_{[S]}$ is given by

$$\hat{\theta}_{[S]} = \hat{\theta} + O_p(N^{-1/2}) = \hat{\theta} - [J_{N,[S]}(\hat{\theta})]^{-1} \partial_{\theta} \log p_S(\hat{\theta}) [1 + O_p(N^{-1/2})], \quad (21)$$

where $J_{N,[S]}(\theta) = -\partial_{\theta}^2 \log p(\theta | Y_{[S]})$. If $\gamma = 0$, then

$$\hat{\theta}_{[S]} = \hat{\theta} - [J_N(\hat{\theta})]^{-1} \partial_{\theta} \log p_S(\hat{\theta}) [1 + O_p(N^{-1/2}) + O_p(N_S/N)]. \quad (22)$$

- b. The one-step approximation for $\tilde{\theta}_{[S]}$ is given by

$$\tilde{\theta}_{[S]} - \tilde{\theta} = (\hat{\theta}_{[S]} - \hat{\theta}) [1 + o_p(1)].$$

- c. CP(S) and CM(S) can be approximated by

$$AP_2(S; \tilde{\theta}) = [\partial_{\theta} \log p_S(\tilde{\theta})]^T [J_{N,[S]}(\tilde{\theta})]^{-1} [J_N(\tilde{\theta})] [J_{N,[S]}(\tilde{\theta})]^{-1} [\partial_{\theta} \log p_S(\tilde{\theta})].$$

If $\gamma = 0$, then

$$CP(S) = CM(S) [1 + o_p(1)] = AP_2(S; \tilde{\theta}) [1 + o_p(1)]. \quad (23)$$

- d. $D_{\phi}(S)$ can be approximated by

$$D_{\phi}(S) = \phi(A_S) + O_p(N^{-1}), \quad (24)$$

where $A_S = \sigma \times p(Y_{[S]} | \hat{\theta}) p(\hat{\theta}) / [\sigma_{[S]} \times p(Y_{[S]} | \hat{\theta}_{[S]}) p(\hat{\theta}_S)]$, $\sigma^2 = |J_N(\hat{\theta})| / N^{-1}$ and $\sigma_{[S]}^2 = |J_{N,[S]}(\hat{\theta}_S)| / N^{-1}$.

Theorem 2 has several important implications. Theorem 2 (a) and (b) provide the one-step approximations of $\hat{\theta}_{[S]}$ and $\tilde{\theta}_{[S]}$, which reduce the burden of computing $\hat{\theta}_{[S]}$ and $\tilde{\theta}_{[S]}$ for each S . Theorem 2 (c) provides the theoretical approximations of CP(S) and CM(S). If $N_S/N \rightarrow 0$, such as $N_S = \sqrt{N}$, then CP(S) and CM(S) can be well approximated by $AP_2(S; \tilde{\theta})$. Theorem 2 (d) shows that when $N_S \rightarrow \infty$, $D_{\phi}(S)$, which can be approximated by $\phi(A_S)$, is not asymptotically equivalent to $AP_2(S; \tilde{\theta})$ in any case. Therefore, we cannot use $AP_2(S; \tilde{\theta})$ to characterize the asymptotic behavior of $D_{\phi}(S)$. Since calculating CM(S), CP(S), $D_{\phi}(S)$ and $p(Y_{[S]} | \theta)$ can be computationally tedious for models with missing data, we generally suggest using their first-order approximations $AP_1(S; \tilde{\theta})$ and $AP_2(S; \tilde{\theta})$ for identifying influential observations. Moreover, we can easily develop a similar procedure for computing $AP_2(S; \tilde{\theta})$.

3 Illustrative Examples

In this section, we illustrate our methodology with simulated data and a real dataset.

3.1 Simulated Data

The goal of our simulations was to evaluate the accuracy of the first-order approximations of the three Bayesian case influence measures in small and moderate sample sizes. We generated 100 data sets from a Binomial mixed model, which has been extensively studied in the literature (Molenberghs and Verbeke, 2005). Specifically, each data set contains n clusters. For each cluster, the random effect b_i , which can be regarded as “missing data”, was first independently generated from a $\mathcal{N}(0, \sigma^2)$ distribution and then, given b_i , the observations $y_{ij} (j = 1, 2, 3; i = 1, \dots, n)$ were independently generated from a Binomial random generator such that $y_{ij} \sim B(n_{ij}, p_{ij})$ with p_{ij} satisfying $\log(p_{ij}/(1 - p_{ij})) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + b_i$, in which the n_{ij} were randomly drawn from $\{1, \dots, 5\}$. Moreover, the covariates \mathbf{x}_{ij} were set as $(1, u_{ij} - 0.5)^T$, and the u_{ij} were independently generated from a $U[0, 1]$ distribution. For all 100 data sets, both the responses and covariates were repeatedly generated, while the true value of $(\boldsymbol{\beta}^T, \sigma^2)$ was fixed at $(0.5, 0.5, 0.5)$. The sample size n was set at 10 and 50, respectively, to represent small and moderate sample sizes.

For each simulated data set, we created two types of influential observations in order to compare the accuracy of the first-order approximations and their capability in the identification of these influential clusters. For the first type, we deleted all the observations in clusters $n - 1$ and n and then reset $\{n_{ij} = 3 : j = 1, 2, 3; i = n - 1, n\}$ and $(b_{n-1}, b_n) = (4, -4)$ to generate y_{ij} for $i = n - 1, n$ and $j = 1, 2, 3$ according to the above binomial random effects model. Thus, the new $(n - 1)$ th and n th clusters can be regarded as influential clusters due to the extreme values of b_{n-1} and b_n . Moreover, the number of observations in these two clusters is relatively small compared to the total number of observations.

For the second type of influential observations, we deleted the observations from the n -th cluster and then reset $\{n_{ij} = 10 : j = 1, 2, 3; i = n\}$ and then generated $\{y_{nj} : j = 1, 2\}$ with $b_n = 0$ and y_{n3} with $b_n = -6$ from the same model. Since different b_n values were used to generate different observations in the n -th cluster, the n -th cluster can be treated as an “outlier”. Moreover, the number of observations in the last cluster is relatively large compared to the total number of observations when $n = 10$.

For each data set, we deleted each cluster one at a time and then calculated the differences between the three Bayesian diagnostic measures and their first-order approximations for each cluster. Since b_i is one-dimensional, we used a standard numerical method to compute $p_S(\boldsymbol{\theta}) = P(\mathbf{Y}_S | \mathbf{Y}_{[S]^c}, \boldsymbol{\theta})$ and to approximate the exact values of the three Bayesian diagnostic measures based on (5) and (6). Moreover, for the true ‘good’ and influential clusters, we computed the average biases and standard errors of these differences (Tables 1-3). For the true ‘good’ clusters, the use of $AP_2(S; \tilde{\boldsymbol{\theta}})$ leads to smaller average biases and comparable standard errors to $CP(S)$ and $CW(S)$ compared with $AP_1(S; \tilde{\boldsymbol{\theta}})$. Increasing the sample size decreases the average bias and standard error of the first-order approximations. Compared with $AP_2(S; \tilde{\boldsymbol{\theta}})$, $AP_1(S; \tilde{\boldsymbol{\theta}})$ is a better approximation to $D_\phi(S)$.

We also computed the frequency of correctly ranking the true influential/outlying clusters as the top four clusters based on the 100 simulated data sets for each scenario (Table 4). Due to the randomness introduced by random number generator, it is possible that few ‘good’ observations may appear as influential observations. Among all three case influence measures, $D_\phi(S)$ is more sensitive for detecting influential/outlying clusters compared to $CP(S)$ and $CW(S)$. Although $AP_1(S; \tilde{\boldsymbol{\theta}})$ and $AP_2(S; \tilde{\boldsymbol{\theta}})$ are not very accurate approximations to the three Bayesian case influence measures (Tables 1-3), they consistently selected the influential clusters as the top four clusters. Furthermore, the results in Table 4 actually indicate that compared with all three Bayesian case influence measures, $AP_1(S; \tilde{\boldsymbol{\theta}})$ and $AP_2(S; \tilde{\boldsymbol{\theta}})$ may be more effective in detecting influential/outlying clusters in all scenarios.

Thus, for the purpose of detecting influential/outlying clusters, the performance of the first-order approximations $AP_1(\mathcal{S}; \hat{\theta})$ and $AP_2(\mathcal{S}; \hat{\theta})$ are quite satisfactory.

Finally, we generated a data set by using the same Binomial mixed model. We deleted all observations of the 38-th cluster with the smallest leverage (Figure 1) and then regenerated $\{y_{38,j}: j = 1, 2, 3\}$ by using the same model except that $x_{38,j,2}$ was changed into $x_{38,j,2} + 5.0$. However, when we fitted the same Binomial mixed model, we used the original $x_{38,j}$. Thus, the 38-th cluster can be regarded as an ‘outlying’ cluster since different $x_{38,j}$ values were used to generate observations. Index plots of $CM(i)$ and $D_{\phi}(i)$ (Figure 2) indicate that $CM(i)$ seems to be more sensitive to outliers with low leverage covariates.

3.2 AIDS Data

We considered a small portion of a real dataset from a study of the relationship between acquired immune deficiency syndrome (AIDS) and the use of condoms (Morisky et al., 1998). This data set contained 11 items on knowledge about AIDS and beliefs, behaviors and attitudes towards condoms use collected from $n = 1116$ subjects. Nine of them were taken as responses in $y_i = (y_{i1}, \dots, y_{i9})^T$ and a continuous item x_{i1} (item 37) and an ordered categorical item x_{i2} (item 21, which was treated as continuous) were taken as covariates. The definitions of the nine items are given in the Appendix. In this dataset, the variables $y_{i1}, y_{i2}, y_{i3}, y_{i7}, y_{i8}$ and y_{i9} were measured via a 5-point scale and hence were treated as continuous; the variables y_{i4}, y_{i5} and y_{i6} were continuous. The responses and covariates are missing at least once for 361 subjects (32%), while the covariate x_{i2} is completely observed. The missing data patterns for the response variables are given in Table 4 of Lee and Tang (2006).

To fit the AIDS data, we considered a complex structural equations modeling in the presence of not missing at random (NMAR) responses and missing at random (MAR) covariates. The responses (y_{i1}, y_{i2}, y_{i3}) are related to a latent variable, η_i which can be interpreted as the ‘threat of AIDS’, while the responses (y_{i4}, y_{i5}, y_{i6}) and (y_{i7}, y_{i8}, y_{i9}) are respectively related to the latent variables ξ_{i1} and ξ_{i2} , which can be respectively interpreted as ‘aggressiveness of the sex worker’ and ‘worry of contracting AIDS’. Specifically, to identify the relationship between the responses y_i and the latent variables $\omega_i = (\eta_i, \xi_{i1}, \xi_{i2})^T$, we consider the following measurement equation:

$$y_i = \mu + \Lambda \omega_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (25)$$

where $\mu = (\mu_1, \dots, \mu_9)^T$ is a vector of intercepts, (ξ_{i1}, ξ_{i2}) is independent of the measurement error vector ε_i , and $(\xi_{i1}, \xi_{i2}) \sim N(\mathbf{0}, \Phi)$ and $\varepsilon_i \sim N(\mathbf{0}, \Psi)$, in which $\Psi = \text{diag}(\psi_1, \dots, \psi_9)$ and $\Phi = (\phi_{ij})$ is a 2×2 covariance matrix. We also assume the following structure for Λ :

$$\Lambda^T = \begin{pmatrix} 1.0* & \lambda_{21} & \lambda_{31} & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* \\ 0.0* & 0.0* & 0.0* & 1.0* & \lambda_{52} & \lambda_{62} & 0.0* & 0.0* & 0.0* \\ 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 1.0* & \lambda_{83} & \lambda_{93} \end{pmatrix}, \quad (26)$$

where 0.0* and 1.0* are regarded as fixed values to identify the scale of the latent factor. To study the relationship between η and (x_1, x_2, ξ_1, ξ_2) , we consider the following nonlinear structural equations model:

$$\eta_i = b_1 x_{i1} + b_2 x_{i2} + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1} \xi_{i2} + \delta_i, \quad (27)$$

where $\delta_i \sim \mathcal{N}(0, \psi_\delta)$. We let $r_{yij} = 1$ if y_{ij} is missing and $r_{yij} = 0$ if y_{ij} is observed; $r_{x_{i1}} = 1$ if x_{i1} is missing and $r_{x_{i1}} = 0$ if x_{i1} is observed. Based on the missingness patterns, we assume that the missing data for the responses is NMAR whilst the missing data for the covariates is MAR. In particular, we consider the following missing data mechanism for y_{ij} :

$$\text{logit} \{ \Pr(r_{yij}=1|\varphi) \} = \varphi_0 + \varphi_1 y_{i1} + \dots + \varphi_9 y_{i9}, \quad (28)$$

where $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \dots, \varphi_9)^T$. Since x_{i1} may be missing, we have to specify a distribution for it. For simplicity, we assume that $x_{i1} \sim \mathcal{N}(0, \psi_x)$. Posterior inference for the structural parameters is obtained via Markov chain Monte Carlo methods as introduced in (Lee and Tang, 2006).

Due to the computational complexity in calculating the exact values of the three Bayesian influence measures, we only calculated the case influence measures $AP_1(\{i\})$ and $AP_2(\{i\})$ (Figure 1). Both case influence measures identify subjects 14, 25, 28, 137, 168, 175, 274, 408, 492, and 985 as influential cases (Figs. 3 and 4). Among them, cases 14 and 985 stand out as the most influential cases. Inspecting Table 5 reveals that case 14 has the largest number of vaginal sex (y_4) in the last 7 days and case 985 has the largest number of blow jobs (y_5) in the last 7 days (Fig. 4 (a) and (c)). Cases 25, 28, 137, and 408 have the largest numbers of hand jobs (y_6) given in the last 7 days, cases 168, 175, and 274 have relatively large numbers of blow jobs compared with other subjects, and case 492 has the second largest number of vaginal sex in the last 7 days (Table 5).

Furthermore, following the classical linear model, we also calculated $h_i = \mathbf{x}_i^T \left(\sum_k \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i$ for all of the fully observed \mathbf{x}_i 's, where \sum_k sums over all fully observed covariates. Among all 10 influential observations, only subject 14 has a high leverage point with $h_{14} = 0.0155$. Inspecting all leverage values reveals that the 93rd subject, who has an extremely high leverage point, has the longest period as a prostitute among all subjects, whereas this subject is not an influential subject.

We further computed the posterior means and standard deviations of all the parameters with and without these 10 influential observations. We observed some effects of deleting these 10 influential observations on the parameters associated with y_4 , y_5 , and y_6 (Table 6). For instance, after deleting the influential observations, λ_{62} was changed from 0.855 to 1.296 and φ_5 was changed from -0.413 to -0.297. This indicates that it is important to identify influential observations and assess their effects on the statistical inference in complex statistical models with missing data.

4 Discussion

We have derived two first-order approximations to three Bayesian case influence measures under the deletion of a small (or large) number of observations. We have shown that the first-order approximation measures are useful tools for detecting influential observations in the presence of missing data for a large class of statistical models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Abraham B, Box GEP. Bayesian analysis of some outlier problems in time series. *Biometrika*. 1979; 66:229–236.
- Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*. 1993; 88:669–679.
- Blyth S. Local divergence and association. *Biometrika*. 1994; 81:579–584.
- Box GEP, Tiao GC. A Bayesian approach to some outlier problems. *Biometrika*. 1968; 55:119–129. [PubMed: 5661037]
- Bradlow ET, Zaslavsky AM. Case influence analysis in Bayesian inference. *Journal of Computational and Graphical Statistics*. 1997; 6:314–331.
- Carlin BP, Polson NG. An expected utility approach to influence diagnostics. *Journal of the American Statistical Association*. 1991; 86:1013–1021.
- Chaloner K. Bayesian residual analysis in the presence of censoring. *Biometrika*. 1991; 78:637–644.
- Christensen, R. *Log-linear Models and Logistic Regression*. Springer-Verlag Inc.; 1997.
- Christensen R, Pearson LM, Johnson W. Case-deletion diagnostics for mixed models. *Technometrics*. 1992; 34:38–45.
- Cook RD. Detection of influential observation in linear regression. *Technometrics*. 1977; 19:15–18.
- Cook, RD.; Weisberg, S. *Residuals and Influence in Regression*. Chapman & Hall; 1982.
- Csiszár I. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*. 1967; 2:299–318.
- Daniels, MJ.; Hogan, JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. London: Chapman and Hall; 2008.
- Epifani I, MacEachern SN, Peruggia M. Case-deletion importance sampling estimators: central limit theorems and related results. *Electronic Journal of Statistics*. 2008; 2:774–806.
- Geisser S. The predictive sample reuse method with applications. *Journal of the American Statistical Association*. 1975; 70:320–328.
- Geisser, S. *Predictive Inference: An Introduction*. London: Chapman & Hall; 1993.
- Gelfand AE, Dey DK. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society Series B*. 1994; 56:501–514.
- Gelfand, AE.; Dey, DK.; Chang, H. Model determination using predictive distributions, with implementation via sampling-based methods (Disc: P160-167). In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. *Bayesian Statistics 4*. Oxford: Oxford University Press; 1992. p. 147-159.
- Ibrahim JG, Chen MH, Lipsitz SR, Herring A. Missing-data methods for generalized linear models: a comparative review. *Journal of the American Statistical Association*. 2005; 100:332–346.
- Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. *Test*. 2009; 18:1–43. [PubMed: 21218187]
- Ibrahim JG, Zhu HT, Tang NS. *Bayesian local influence for survival models*. Lifetime Data Analysis. 2010
- Johnson W. Influence measures for logistic regression: another point of view. *Biometrika*. 1985; 72:59–65.
- Johnson W, Geisser S. A predictive view of the detection and characterization of influential observations in regression analysis. *Journal of the American Statistical Association*. 1983; 78:137–144.
- Kass, RE.; Kadane, JB.; Tierney, L. Asymptotic evaluation of integrals arising in Bayesian inference. In: Page, C.; LePage, R., editors. *Computing Science and Statistics: Proceedings of the Symposium on the Interface*. Springer-Verlag Inc.; 1990. p. 38-42.
- Kass RE, Tierney L, Kadane JB. Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika*. 1989; 76:663–674.
- Lee, SY. *Structural Equation Modelling: A Bayesian Approach*. Wiley; 2007.
- Lee SY, Tang NS. Analysis of nonlinear structural equation models with nonignorable missing covariates and ordered categorical data. *Statistica Sinica*. 2006; 16:1117–1141.

- Little RJA. Regression with missing x's: A review. *Journal of the American Statistical Association*. 1992; 87:1227–1237.
- Little, RJA.; Rubin, DB. *Statistical Analysis With Missing Data*. New York: Wiley; 2002.
- Louis TA. Finding the observed information using the em algorithm. *Journal of the Royal Statistical Society Series B*. 1982; 44:226–233.
- McCulloch RE. Local model influence. *Journal of the American Statistical Association*. 1989; 84:473–478.
- Molenberghs, G.; Kenward, G. *Missing Data in Clinical Studies*. Wiley; New York: 2007.
- Molenberghs, G.; Verbeke, G. *Models for Discrete Longitudinal Data*. Springer; New York: 2005.
- Morisky DE, Tiglaio TV, Sneed CD, Tempongko SB, Baltazar JC, Detels R, Stein JA. The effects of establishment practices, knowledge and attitudes on condom use among filipina sex workers. *AIDS Care*. 1998; 10:213–320. [PubMed: 9625904]
- Peng F, Dey DK. Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*. 1995; 23:199–213.
- Peruggia M. On the variability of case-deletion importance sampling weights in the bayesian linear model. *Journal of the American Statistical Association*. 1997; 92:199–207.
- Pettit LI. Diagnostics in Bayesian model choice. *The Statistician: Journal of the Institute of Statisticians*. 1986; 35:183–190.
- Skrondal, A.; Rabe-Hesketh, S. *Generalized Latent Variable Modeling*. London: Chapman and Hall; 2004.
- Tierney L, Kass RE, Kadane JB. Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*. 1989; 84:710–716.
- Wang HM, Jones MP, Burns TL. Regression diagnostics for the class a regressive model with quantitative phenotypes. *Genetic Epidemiology*. 1999; 17:174–187. [PubMed: 10446465]
- Wei, B. *Exponential Family Nonlinear Models*. Singapore: Springer-Verlag; 1998.
- Weiss R. An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society Series B*. 1996; 58:739–750.
- Weiss RE, Cook RD. A graphical case statistic for assessing posterior influence. *Biometrika*. 1992; 79:51–55.
- Zhu H, Lee SY, Wei BC, Zhou J. Case deletion measures for models with incomplete data. *Biometrika*. 2001; 88:727–737.
- Zhu, HT.; Ibrahim, JG.; Cho, N.; Tang, NS. A general review of Bayesian influence analysis. In: Chen, MH.; Dey, DK.; Muller, P.; Sun, D.; Ye, K., editors. *Frontier of statistical decision making and Bayesian analysis*. New York: Springer-Verlag; 2010. p. 219-237.

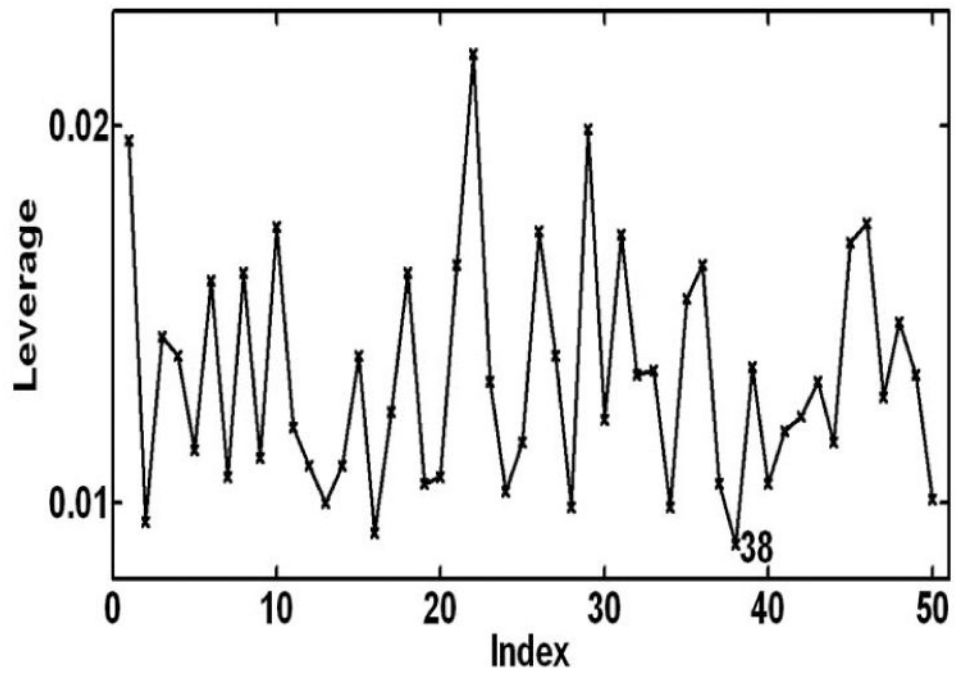


Figure 1.
Index plot of leverage values in simulation study.

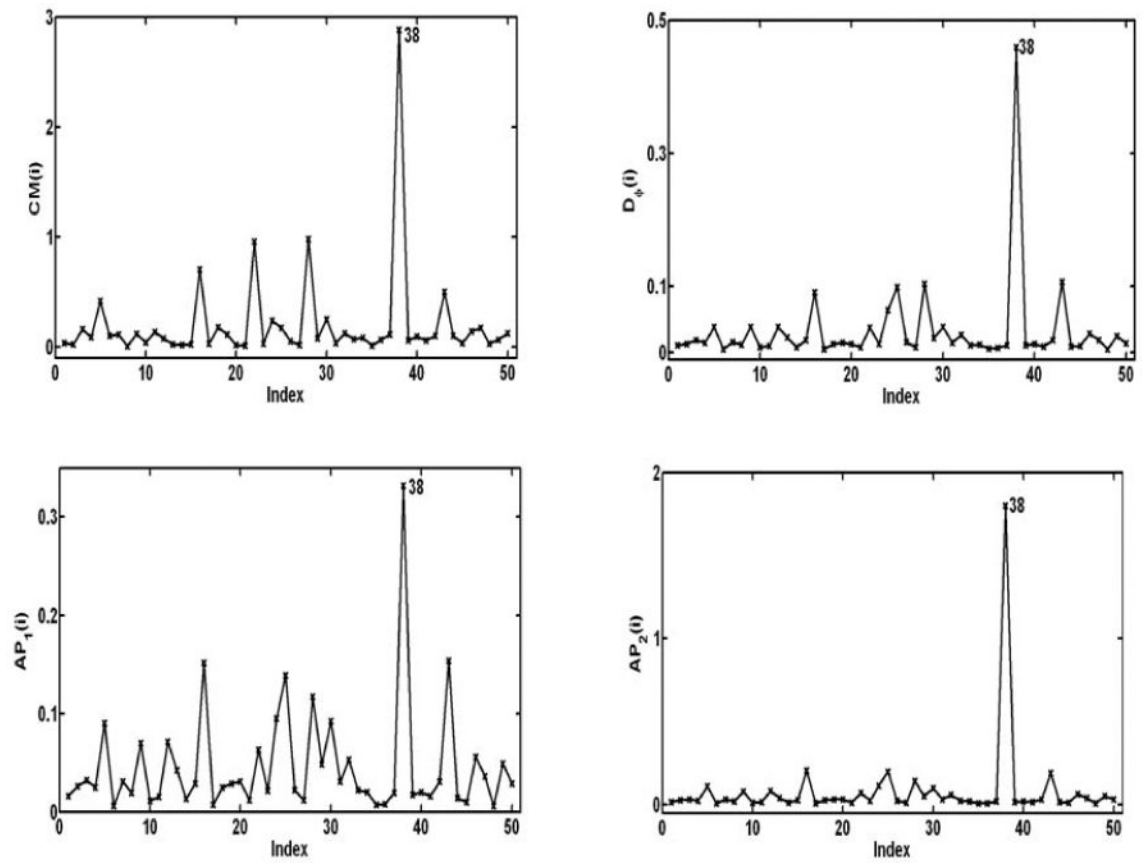


Figure 2. Index plots of $CM(i)$, $D_\phi(i)$, $AP_1(i)$ and $AP_2(i)$ in simulation study.

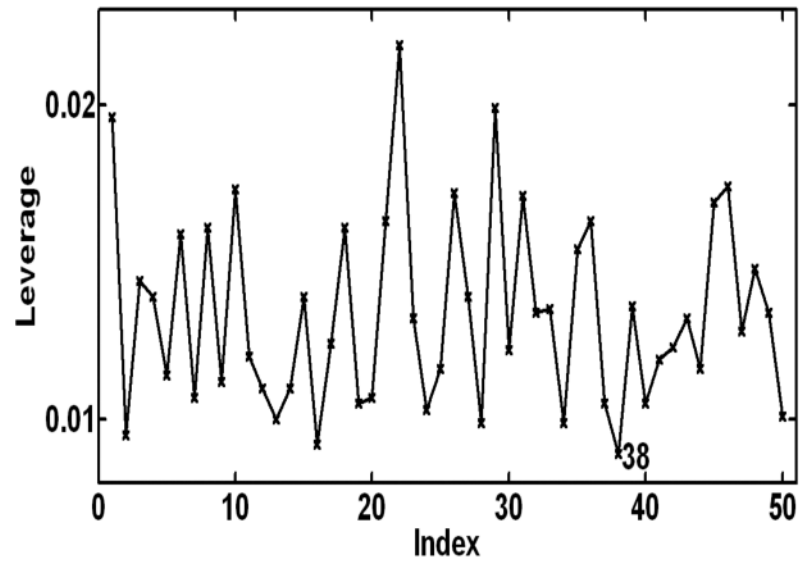


Figure 3. AIDS data: index plots of diagnostic measures (a) $AP_1(i)$ and (b) $AP_2(i)$.

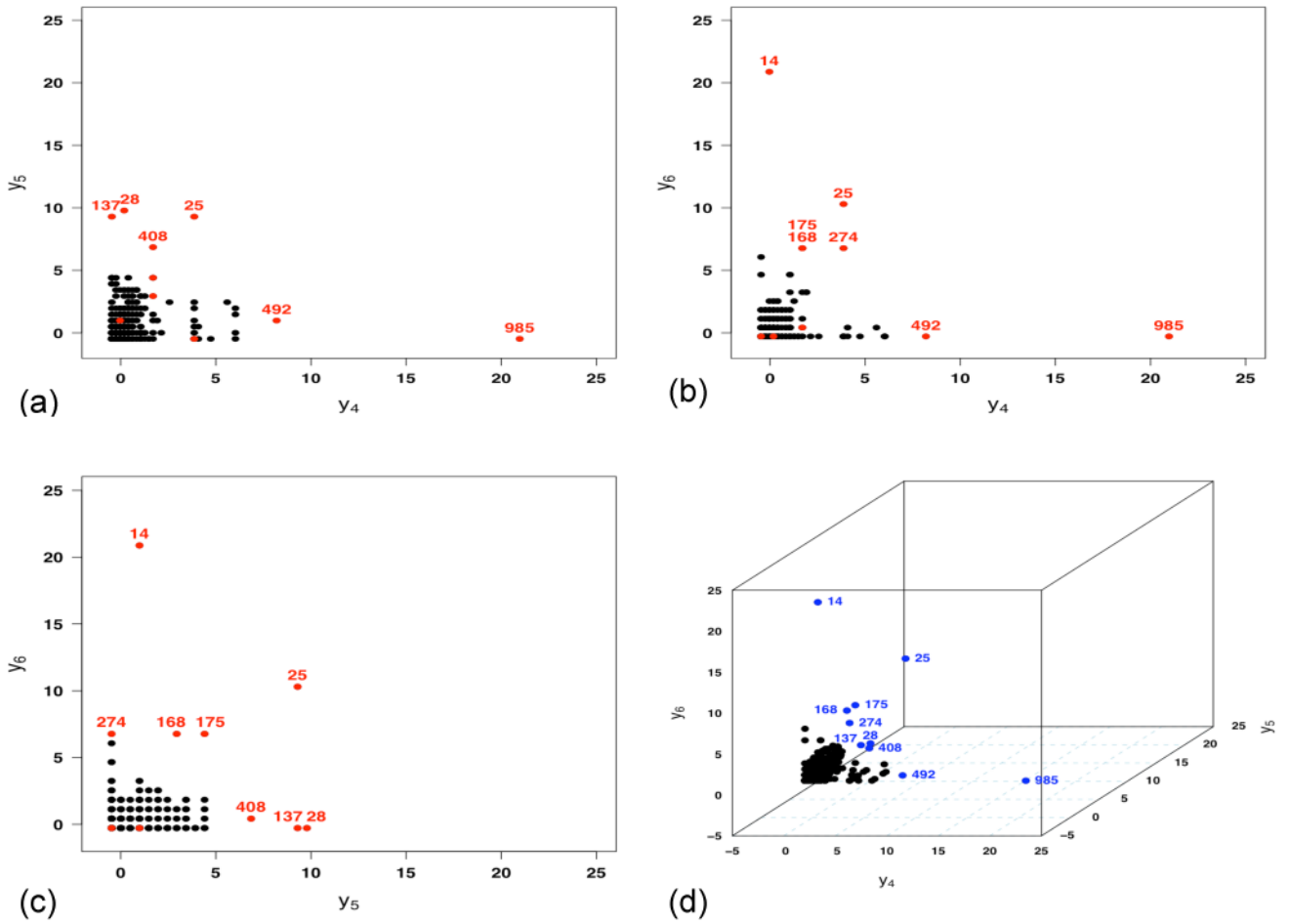


Figure 4. AIDS data: scatter plots of (a) (y_4, y_5) , (b) (y_4, y_6) , (c) (y_5, y_6) , and (d) (y_4, y_5, y_6) .

Results from simulation studies for two types of influential observations for $n = 10$. Average biases (ABs) and standard errors (SEs) of the differences between the three Bayesian case influence measures and their first-order approximations. For the first type, clusters 9 and 10 are influential clusters, while cluster 10 is the outlying cluster for the second type.

Table 1

Cluster	AP ₁ (AB)			AP ₂ (AB)			AP ₁ (SE)			AP ₂ (SE)			
	D _f	CP	CM	D _f	CP	CM	D _f	CP	CM	D _f	CP	CM	
I	1	-0.055	0.077	0.057	-0.119	0.013	-0.007	0.062	0.133	0.098	0.159	0.075	0.088
	2	-0.050	0.079	0.059	-0.124	0.005	-0.015	0.065	0.165	0.138	0.235	0.170	0.148
	3	-0.057	0.076	0.063	-0.136	-0.003	-0.016	0.074	0.141	0.118	0.211	0.089	0.086
	4	-0.053	0.070	0.055	-0.117	0.006	-0.008	0.064	0.088	0.086	0.147	0.075	0.065
	5	-0.059	0.058	0.045	-0.128	-0.011	-0.025	0.068	0.082	0.092	0.197	0.132	0.139
	6	-0.046	0.055	0.058	-0.101	0.001	0.003	0.047	0.106	0.104	0.119	0.071	0.065
	7	-0.062	0.086	0.062	-0.150	-0.003	-0.027	0.071	0.125	0.124	0.230	0.130	0.152
	8	-0.061	0.074	0.061	-0.137	-0.002	-0.015	0.055	0.135	0.107	0.180	0.113	0.111
	9*	-0.200	0.129	0.080	-0.532	-0.203	-0.252	0.100	0.223	0.201	0.478	0.327	0.355
	10*	-0.348	0.250	0.012	-1.380	-0.782	-1.021	0.108	0.249	0.220	0.913	0.789	0.923
II	1	-0.066	0.081	0.074	-0.151	-0.004	-0.011	0.080	0.137	0.147	0.239	0.180	0.216
	2	-0.050	0.053	0.048	-0.097	0.005	0.000	0.064	0.077	0.076	0.125	0.068	0.087
	3	-0.063	0.061	0.058	-0.132	-0.008	-0.011	0.069	0.097	0.108	0.172	0.098	0.129
	4	-0.073	0.088	0.074	-0.176	-0.016	-0.029	0.092	0.134	0.162	0.262	0.129	0.165
	5	-0.057	0.062	0.057	-0.109	0.010	0.005	0.066	0.103	0.100	0.134	0.085	0.093
	6	-0.069	0.063	0.052	-0.144	-0.013	-0.024	0.078	0.096	0.088	0.198	0.123	0.122
	7	-0.062	0.076	0.050	-0.137	0.000	-0.026	0.078	0.143	0.131	0.202	0.109	0.156
	8	-0.066	0.078	0.113	-0.172	-0.029	0.006	0.077	0.119	0.171	0.324	0.219	0.284
	9	-0.068	0.071	0.074	-0.162	-0.023	-0.020	0.088	0.173	0.186	0.304	0.234	0.269
	10*	0.085	0.681	0.301	-2.637	-2.041	-2.421	0.771	0.544	0.509	2.549	2.926	3.014

Selected results from simulation studies for $n = 50$ and the first type of influential observations. Average biases (ABs) and standard errors (SEs) of the differences between the three Bayesian case influence measures and their first-order approximations. Note that clusters 49 and 50 are influential clusters.

Table 2

Cluster	AP ₁ (AB)			AP ₂ (AB)			AP ₁ (SE)			AP ₂ (SE)		
	D _↓	CP	CM	D _↓	CP	CM	D _↓	CP	CM	D _↓	CP	CM
1	-0.023	0.012	-0.002	-0.028	0.006	-0.008	0.027	0.026	0.026	0.041	0.031	0.038
5	-0.022	0.014	0.003	-0.027	0.009	-0.002	0.025	0.028	0.023	0.035	0.030	0.028
10	-0.018	0.014	0.004	-0.022	0.010	-0.000	0.025	0.023	0.022	0.036	0.027	0.029
15	-0.022	0.013	0.004	-0.026	0.008	-0.001	0.032	0.027	0.021	0.041	0.028	0.025
20	-0.021	0.016	0.007	-0.026	0.011	0.002	0.026	0.029	0.033	0.041	0.033	0.044
25	-0.016	0.011	0.003	-0.019	0.009	0.001	0.016	0.024	0.017	0.020	0.025	0.018
30	-0.022	0.010	0.003	-0.027	0.005	-0.002	0.030	0.029	0.028	0.043	0.030	0.036
35	-0.022	0.012	-0.001	-0.028	0.005	-0.008	0.027	0.023	0.035	0.043	0.029	0.048
40	-0.022	0.011	0.005	-0.027	0.006	0.000	0.024	0.028	0.030	0.033	0.032	0.035
45	-0.023	0.014	0.002	-0.029	0.008	-0.004	0.025	0.026	0.023	0.039	0.029	0.032
49*	-0.051	0.005	-0.017	-0.066	-0.009	-0.032	0.016	0.041	0.037	0.022	0.043	0.039
50*	-0.141	-0.001	-0.137	-0.238	-0.098	-0.235	0.070	0.083	0.101	0.124	0.122	0.154

Selected results from simulation studies for $n = 50$ and the first type of influential observations. Average biases (ABs) and standard errors (SEs) of the differences between the three Bayesian case influence measures and their first-order approximations. Note that cluster 50 is the outlying cluster.

Table 3

Cluster	AP ₁ (AB)			AP ₂ (AB)			AP ₁ (SE)			AP ₂ (SE)		
	D _↓	CP	CM	D _↓	CP	CM	D _↓	CP	CM	D _↓	CP	CM
1	-0.019	0.008	0.010	-0.022	0.005	0.007	0.019	0.024	0.043	0.026	0.026	0.046
5	-0.020	0.011	0.005	-0.023	0.007	0.002	0.021	0.022	0.028	0.029	0.024	0.032
10	-0.016	0.010	0.011	-0.018	0.007	0.008	0.016	0.020	0.034	0.019	0.020	0.034
15	-0.016	0.014	0.010	-0.020	0.010	0.007	0.016	0.025	0.033	0.022	0.023	0.034
20	-0.020	0.012	0.007	-0.024	0.008	0.003	0.018	0.025	0.031	0.025	0.026	0.033
25	-0.016	0.009	0.006	-0.019	0.006	0.004	0.018	0.022	0.028	0.023	0.023	0.030
30	-0.019	0.009	0.014	-0.022	0.005	0.010	0.019	0.028	0.045	0.026	0.030	0.046
35	-0.026	0.008	0.004	-0.032	0.001	-0.002	0.028	0.031	0.047	0.040	0.034	0.057
40	-0.021	0.014	0.011	-0.025	0.010	0.006	0.021	0.028	0.038	0.029	0.030	0.040
45	-0.024	0.007	0.005	-0.031	0.000	-0.002	0.035	0.032	0.041	0.062	0.055	0.065
50*	-0.162	0.034	-0.170	-0.510	-0.314	-0.518	0.058	0.081	0.164	0.416	0.414	0.507

Table 4

Frequency of correctly ranking the influential clusters as the top 4 clusters based on 100 replications for four different scenarios.

<i>n</i>	Type	D_4	CP	CM	AP ₁	AP ₂
10	I	88	84	85	91	91
10	II	99	99	99	99	99
50	I	96	74	94	96	96
50	II	100	91	100	100	100

Table 5

The observed responses and covariates of ten influential observations from the AIDS data set. Note that variables y_4, y_5, y_6 and x_1 are centered. ‘No’ denotes the subject number and missing value is represented as ‘.’.

No	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	x_1	x_2
14	3	3	5	-0.03	0.98	20.88	3	4	3	4.12	1
25	4	4	3	3.86	9.29	10.30	5	1	4	-0.14	3
28	1	2	2	0.18	9.78	-0.28	3	1	3	-0.25	2
137	2	3	4	-0.46	9.29	-0.28	5	2	2	0.93	2
168	3	3	2	1.69	2.94	6.77	3	2	3	-0.33	4
175	4	4	.	1.69	4.40	6.77	5	1	.	0.93	4
274	5	3	5	3.86	-0.48	6.77	5	1	5	-0.37	3
408	2	5	5	1.69	6.85	0.42	5	.	5	1.84	5
492	2	5	2	8.19	0.98	-0.28	2	5	5	0.93	4
985	3	1	3	20.97	-0.48	-0.28	5	5	3	-0.45	1

Table 6

AIDS data: posterior means and standard deviations of the parameters with and without 10 influential observations. ES denotes the posterior mean and SD denotes the posterior standard deviation. Parameters with relatively large changes are highlighted.

Results without 10 influential observations													
	μ_1	μ_2	μ_3	μ_4^*	μ_5	μ_6	μ_7	μ_8	μ_9				
ES	0.213	0.164	0.258	-0.042	-0.041	-0.047	0.159	0.120	0.229				
SD	0.025	0.030	0.024	0.021	0.025	0.019	0.023	0.028	0.028				
	λ_{21}	λ_{31}	λ_{52}	λ_{62}^*	λ_{83}	λ_{93}	ψ_6	ψ_x					
ES	0.340	0.579	1.957	0.855	0.691	1.176	0.230	1.314					
SD	0.093	0.101	0.305	0.126	0.160	0.272	0.035	0.861					
	ψ_1	ψ_2	ψ_3	ψ_4^*	ψ_5^*	ψ_6^*	ψ_7	ψ_8	ψ_9				
ES	0.348	0.802	0.398	0.423	0.399	0.336	0.430	0.770	0.474				
SD	0.038	0.039	0.025	0.022	0.055	0.019	0.030	0.036	0.039				
	b_1	b_2	γ_1	γ_2	γ_3	ϕ_{11}	ϕ_{12}	ϕ_{22}	ϕ_0				
ES	-0.047	0.059	-0.233	0.290	-0.123	0.079	-0.014	0.099	-2.976				
SD	0.024	0.022	0.098	0.133	0.217	0.015	0.006	0.025	0.073				
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5^*	ϕ_6	ϕ_7	ϕ_8	ϕ_9				
ES	0.037	-0.263	0.242	-0.218	-0.413	-0.170	-0.383	-0.071	0.171				
SD	0.083	0.077	0.124	0.103	0.092	0.097	0.077	0.075	0.130				
Results without 10 influential observations													
	μ_1	μ_2	μ_3	μ_4^*	μ_5	μ_6	μ_7	μ_8	μ_9				
ES	0.209	0.171	0.259	-0.003	0.003	0.002	0.158	0.116	0.234				
SD	0.026	0.030	0.023	0.029	0.031	0.030	0.023	0.029	0.029				
	λ_{21}	λ_{31}	λ_{52}	λ_{62}^*	λ_{83}	λ_{93}	ψ_6	ψ_x					
ES	0.315	0.584	2.126	1.296	0.686	1.198	0.227	1.326					
SD	0.102	0.082	0.286	0.145	0.187	0.246	0.030	0.889					

	ψ_1	ψ_2	ψ_3	ψ_4^*	ψ_5^*	ψ_6^*	ψ_7	ψ_8	ψ_9
ES	0.348	0.798	0.399	0.890	0.551	0.825	0.437	0.774	0.472
SD	0.031	0.039	0.024	0.041	0.074	0.049	0.029	0.037	0.046
	b_1	b_2	γ_1	γ_2	γ_3	ϕ_{11}	ϕ_{12}	ϕ_{22}	ϕ_0
ES	-0.039	0.057	-0.283	0.333	-0.041	0.100	-0.016	0.096	-2.961
SD	0.023	0.023	0.112	0.104	0.200	0.015	0.009	0.022	0.071
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5^*	ϕ_6	ϕ_7	ϕ_8	ϕ_9
ES	0.042	-0.245	0.288	-0.251	-0.297	-0.097	-0.372	-0.091	0.205
SD	0.085	0.074	0.120	0.103	0.077	0.071	0.076	0.073	0.139