Biometrics 65, 618–625
June 2009

DOI: 10.1111/j.1541-0420.2008.01105.x

Median Regression Models for Longitudinal Data with Dropouts

Grace Y. Yi

Department of Statistics and Actuarial Science, University of Waterloo,
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
email: yyi@uwaterloo.ca

and

Wenqing He

Department of Statistical and Actuarial Sciences, University of Western Ontario,
1151 Richmond Street North, London, Ontario N6A 5B7, Canada

Summary. Recently, median regression models have received increasing attention. When continuous responses follow a distribution that is quite different from a normal distribution, usual mean regression models may fail to produce efficient estimators whereas median regression models may perform satisfactorily. In this article, we discuss using median regression models to deal with longitudinal data with dropouts. Weighted estimating equations are proposed to estimate the median regression parameters for incomplete longitudinal data, where the weights are determined by modeling the dropout process. Consistency and the asymptotic distribution of the resultant estimators are established. The proposed method is used to analyze a longitudinal data set arising from a controlled trial of HIV disease (Volberding et al., 1990, *The New England Journal of Medicine* **322,** 941–949). Simulation studies are conducted to assess the performance of the proposed method under various situations. An extension to estimation of the association parameters is outlined.

Key words: Logistic regression models; Longitudinal data; Median regression model; Missing data; Weighted estimating equations.

1. Introduction

Inverse probability weighted generalized estimating equations (IPWGEE) are proposed by Robins, Rotnitzky, and Zhao (1995) to deal with incomplete longitudinal data arising from a missing at random (MAR) mechanism. Since then there has been considerable research on this marginal method (e.g, Fitzmaurice, Molenberghs, and Lipsitz, 1995; Yi and Cook, 2002a, 2002b). This approach is widely viewed as attractive because it does not require complete specification of the joint distribution of the longitudinal responses but rather is based only on specification of the first two moments. In the formulation of IPWGEE, mean regression models are typically utilized to modulate marginal responses.

Mean regression models are feasible for modeling data whose distribution is normal or nearly normal. When distributions have heavy tails or are highly skewed, usual mean regression models may fail to produce efficient estimates; but median regression models would perform satisfactorily (Koenker and Bassett, 1978; Koenker and Xiao, 2001). For a Cauchy error distribution, for instance, the mean regression estimate has an infinite variance, whereas the median regression estimate just has a finite variance. Recently, there has been increasing interest in median regression models. For instance, in survival analysis median regression models have been discussed extensively. Ying, Jung, and Wei (1995) advocated that the median is a simple and meaningful measure for the center of a long-tailed survival distribution and it can be better estimated than the mean in settings with moderate censoring. Yang (1999) and Portnoy (2003) developed methods for median regression by forming a weighted form of hazard and survival functions. Bang and Tsiatis (2002) proposed semiparametric procedures for estimating the parameters in median regression models for censored data. In the context of generalized linear models, Morgenthaler (1992) discussed median regression models based on the quasilikelihood approach. Jung (1996) explored median regression models to address dependent observations. He, Fu, and Fung (2003) and Hogan and Lee (2004) discussed quantile models for complete longitudinal data.

Although longitudinal studies are frequently designed to collect data on every individual in the sample at each assessment, incomplete data often arise. Subjects may drop out of the study before the end of the follow-up. In many contexts the distribution of data may be far from a normal or symmetrical one. To handle data with these features, in this article we explore using median regression models to analyze longitudinal data with dropouts and propose a weighted estimating equation approach. The proposed method is essentially different from that of Robins et al. (1995) as the proposed weighted estimating functions are discontinuous for parameters, and

therefore, the asymptotic properties established in Robins et al. (1995) for the mean regression estimators are not applicable to the current situation (Newey and McFadden, 1994). In this article, we establish the asymptotic properties for the resultant estimators.

This article is partially motivated by longitudinal data arising from a controlled trial of HIV disease (Volberding et al., 1990). Infection with HIV causes chronic, progressive depletion of CD4+ cells, and this depletion, together with the infection of macrophages and other cells, would create an immune deficiency that leads to the cancers and opportunistic infections characteristic of the AIDS. As one of primary objectives, the study investigates the treatment effect of Zidovudine on increasing CD4+ cell counts. Eight hundred and ninety-two adults were randomized to a treatment or control group. The subjects were followed up longitudinally and the measurements were collected at weeks 8, 16, 32, and 48, respectively. The distributions of the outcomes in treatment and control arms are skewed with a number of outliers. Figure 1 displays the distributions of CD4+ cell counts for treatment and control arms at the initial assessment. Furthermore, there was a portion of the participants who were lost to follow-up during the trial. To address these issues, we develop an analysis method based on median regression models with dropouts accounted for.

The remainder of this article is organized as follows. Section 2 introduces the notation and framework, and in Section 3 we discuss estimation and inference procedures. In Section 4, we analyze the CD4+ cell count data with the proposed method. Simulation studies are conducted to assess the performance of the proposed method under a variety of situations and the results are reported in Section 5. An extension to estimation of the association parameters is outlined in Section 6. General discussion is presented in the last section.

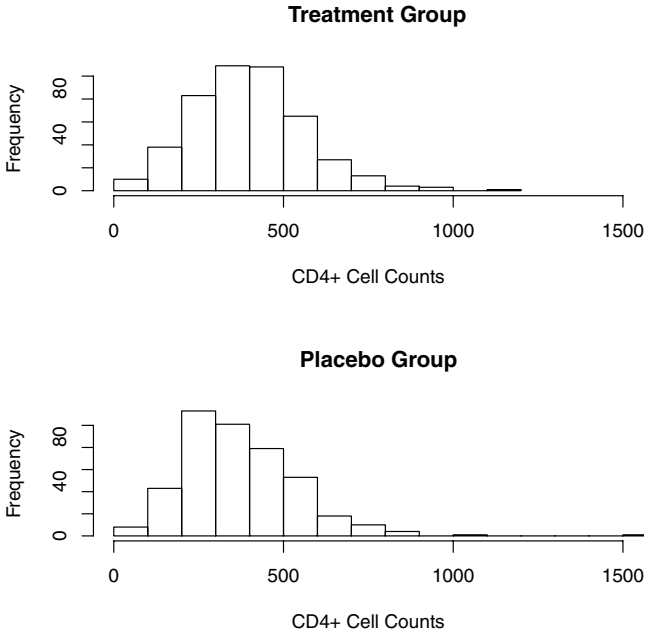

Figure 1. Histograms of the CD4+ cell counts for the treatment and placebo groups at the initial assessment.

2. Notation and Framework

2.1 *Median Regression Models*

Let Y_{ij} be the continuous response for subject i at time point j, and $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijp})'$ be the $p \times 1$ covariate vector for subject i at time j, $j = 1, 2, \ldots, m$; $i = 1, 2, \ldots, n$. Here $\boldsymbol{x}_{ij}$ may solely consist of baseline covariates such as gender, age, and treatment status, or of functions of baseline covariates. It may also consist of external covariates that are time varying in the sense of Robins et al. (1995). Let $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im})'$ and $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}', \boldsymbol{x}_{i2}', \ldots, \boldsymbol{x}_{im}')'$ be the response and covariate vectors for subject i, respectively. Given the covariates $\boldsymbol{x}_i$, let μ_{ij} be the median of Y_{ij}, and $\phi^{-1}f(\mu_{ij})$ be the probability density function of Y_{ij} at μ_{ij}, where $\phi > 0$ is a scalar dispersion parameter, and f is a known function. Here we assume that f is continuous with $f(\mu_{ij}) > 0$. This assumption implies that the cumulative distribution function of Y_{ij} is continuous and differentiable in a neighborhood of μ_{ij}. Furthermore, it warrants the uniqueness of the median. Denote $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{im})'$.

For $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$, consider the median regression model

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}'\boldsymbol{\beta}, \tag{1}$$

where g is a known link function, and $\boldsymbol{\beta}$ is the vector of regression parameters. This model implies that the dependence of median μ_{ij} on the subject level covariates $\boldsymbol{x}_i$ is completely reflected by the time-specific covariates $\boldsymbol{x}_{ij}$. An analogous assumption has been widely used in modeling longitudinal data with mean regression. See Fitzmaurice et al. (1995) and Cook, Zeng, and Yi (2004), for example.

Median regression models have a long history. It has been established that, for instance, sample median is the maximum likelihood estimate of a double exponential distribution. As discussed in Morgenthaler (1992), modeling the median rather than mean is required when using the least-absolute-deviations (LAD) approach. If all response components Y_{ij} are independent, the LAD estimate of $\boldsymbol{\beta}$ is obtained by minimizing $\sum_{i=1}^{n}\sum_{j=1}^{m}|Y_{ij} - \mu_{ij}|$.

2.2 *Models for the Dropouts*

Let R_{ij} be 1 if Y_{ij} is observed, and 0 otherwise. Dropouts or monotone missing data patterns are considered here. That is, $R_{ij} = 0$ implies $R_{ij'} = 0$ for all $j' > j$. Denote $\boldsymbol{R}_i = (R_{i1}, R_{i2}, \ldots, R_{im})'$. Without loss of generality, assume that $R_{i1} = 1$ for every subject i. According to the dependence of the missing data process on the response process, missing data mechanisms may be classified as missing completely at random (MCAR), MAR, and not missing at random (NMAR) (e.g., Kenward, 1998). In this article, we assume an MAR mechanism for the dropout process. Namely, given the covariates, the conditional distribution $f(\boldsymbol{r}_i \,|\, \boldsymbol{x}_i, \boldsymbol{y}_i)$ of the missing data indicator vector $\boldsymbol{R}_i$ depends on the observed response components $\boldsymbol{y}_i^{\text{obs}}$ only.

Let $\lambda_{ij} = P(R_{ij} = 1 \,|\, R_{i,j-1} = 1, \boldsymbol{x}_i, \boldsymbol{y}_i)$, and $\pi_{ij} = P(R_{ij} = 1 \,|\, \boldsymbol{x}_i, \boldsymbol{y}_i)$. Note that $\pi_{ij} = \prod_{t=2}^{j}\lambda_{it}$. Let $H_{ij}^{y} = \{y_{i1}, \ldots, y_{i,j-1}\}$ denote the response history up to (but not

including) time point j. Logistic regression models are commonly used to model the dropout process:

$$\text{logit } \lambda_{ij} = \boldsymbol{u}'_{ij}\boldsymbol{\alpha}, \tag{2}$$

where $\boldsymbol{u}_{ij}$ is the vector consisting of the information of the covariates $\boldsymbol{x}_i$ and the observed responses H_{ij}^y, and $\boldsymbol{\alpha}$ is the vector of regression parameters of dimension q, say. This model implies that the conditional probability λ_{ij} is determined by the response history H_{ij}^y, given the covariates and that subject i is in the study at the previous time point $j-1$. Model (2) has been commonly used in modeling the dropout process. It features an MAR mechanism, but not vice versa.

Let M_i be the random dropout time for subject i, and m_i be a realization, $i = 1, 2, \ldots, n$. Define $L_i(\boldsymbol{\alpha}) = (1-\lambda_{im_i})\prod_{t=2}^{m_i-1}\lambda_{it}$, if $m_i < m$; and $L_i(\boldsymbol{\alpha}) = \prod_{j=2}^{m_i}\lambda_{ij}$, if $m_i = m$, where λ_{it} is determined by model (2). Let $\boldsymbol{S}_i(\boldsymbol{\alpha}) = \partial \log L_i(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}$ be the vector of score functions contributed from subject i. Solving

$$\boldsymbol{S}(\boldsymbol{\alpha}) = \sum_{i=1}^{n}\boldsymbol{S}_i(\boldsymbol{\alpha}) = \boldsymbol{0}, \tag{3}$$

leads to the estimator $\hat{\boldsymbol{\alpha}}$.

3. Estimation and Inference

3.1 *Estimating Equations for* $\boldsymbol{\beta}$

For $i = 1, 2, \ldots, n$, denote $\boldsymbol{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}'$. Let $\boldsymbol{\Gamma}_i = \text{diag}(f(\mu_{ij}),\ j = 1, 2, \ldots, m)$, $\epsilon_{ij} = I(Y_{ij} \geqslant \mu_{ij}) - 1/2$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{im})'$. Here $I(.)$ is the indicator function. Denote $p_{ijk} = P(Y_{ij} \geqslant \mu_{ij}, Y_{ik} \geqslant \mu_{ik} \mid \boldsymbol{x}_i)$ for $j \neq k$. It is easily seen that the covariance matrix of $\boldsymbol{\epsilon}_i$ is given by $\boldsymbol{V}_i = \text{var}(\boldsymbol{\epsilon}_i) = [v_{ijk}]_{m\times m}$, where $v_{ijj} = 1/4$, and $v_{ijk} = p_{ijk} - 1/4$ for $j \neq k$.

When there are no missing data, the generalized estimating equations for $\boldsymbol{\beta}$ are given by

$$\sum_{i=1}^{n}\phi^{-1}\boldsymbol{D}_i'\boldsymbol{\Gamma}_i\boldsymbol{V}_i^{-1}\boldsymbol{\epsilon}_i = \boldsymbol{0}. \tag{4}$$

This formulation is discussed by Jung (1996) by means of the quasilikelihood method. It has been shown that the resulting estimators have minimal asymptotic variance within a certain class of consistent estimators. Godambe (2001) justified a similar form from the viewpoint of optimum estimating functions.

If missing observations are present, then estimating functions in equation (4) constructed from the observed outcomes are no longer unbiased if missing data mechanisms are not MCAR. A proper adjustment accommodating the dropout information is needed. Let $\boldsymbol{\Delta}_i = \text{diag}(I(R_{ij} = 1)/\pi_{ij},\ j = 1, 2, \ldots, m)$ be the weight matrix accommodating missingness, and $\boldsymbol{U}_i = \phi^{-1}\boldsymbol{D}_i'\boldsymbol{\Gamma}_i\boldsymbol{V}_i^{-1}\boldsymbol{\Delta}_i\boldsymbol{\epsilon}_i$, for $i = 1, 2, \ldots, n$. Then, for a given $\boldsymbol{\alpha}$,

$$\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n}\boldsymbol{U}_i, \tag{5}$$

are unbiased estimating functions for $\boldsymbol{\beta}$, provided model (2) for the missingness probability λ_{ij} is correctly specified. Indeed,

$$\begin{aligned} E(\boldsymbol{U}_i) &= E_{Y|X}E_{R|(Y,X)}\left[\phi^{-1}\boldsymbol{D}_i'\boldsymbol{\Gamma}_i\boldsymbol{V}_i^{-1}\boldsymbol{\Delta}_i\boldsymbol{\epsilon}_i\right] \\ &= E_{Y|X}\left[\phi^{-1}\boldsymbol{D}_i'\boldsymbol{\Gamma}_i\boldsymbol{V}_i^{-1}\right. \\ &\qquad \left.\cdot\,\text{diag}\left(\frac{P(R_{ij}=1\mid\boldsymbol{x}_i,\boldsymbol{y}_i)}{\pi_{ij}}, j=1,2,\ldots,m\right)\cdot\boldsymbol{\epsilon}_i\right] \\ &= \boldsymbol{0}. \end{aligned}$$

We note that the form of $\boldsymbol{U}_i$ in equation (5) is analogous to the structure of the IPWGEE for estimating mean regression parameters discussed in Robins et al. (1995). However, an essential difference exists in the estimating functions for the mean and median regression parameters. The former functions are continuous functions of parameters for which asymptotic properties have been well established (e.g., Robins et al., 1995), whereas the latter ones are discontinuous, and this introduces additional complexity in establishing the theoretical properties of the resultant estimators.

3.2 *Estimation Algorithm*

We note that in practical situations, $\boldsymbol{\alpha}$ (or π_{ij}) is unknown and needs to be replaced with a consistent estimate when using $\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ to estimate $\boldsymbol{\beta}$. In this section, we describe a three-stage estimation procedure. As ϕ is a constant, we can drop ϕ from equation (5) when actually performing estimation of $\boldsymbol{\beta}$. Let $\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = (U_{(1)}(\boldsymbol{\beta}, \boldsymbol{\alpha}), \ldots, U_{(p)}(\boldsymbol{\beta}, \boldsymbol{\alpha}))'$, where $U_{(k)}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ denotes the estimating function for parameter β_k. Given $\boldsymbol{\beta}^{(t)} = (\beta_1^{(t)}, \beta_2^{(t)}, \ldots, \beta_p^{(t)})'$, denote $\boldsymbol{\beta}_{(k)}^{(t+1)} = (\beta_1^{(t+1)}, \ldots, \beta_{k-1}^{(t+1)}, \beta_k, \beta_{k+1}^{(t)}, \ldots, \beta_p^{(t)})'$ for $k = 1, 2, \ldots, p$, and $t = 0, 1, \ldots$.

Stage 1: Estimation of $\boldsymbol{\alpha}$

Using the Newton–Raphson algorithm (e.g., Press et al., 1992, Chapter 9) we solve $\boldsymbol{\alpha}$ from equation (3). That is, given an initial value $\boldsymbol{\alpha}^{(0)}$, update $\boldsymbol{\alpha}$ iteratively by

$$\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}^{(t-1)} - \left[\frac{\partial\boldsymbol{S}(\boldsymbol{\alpha}^{(t-1)})}{\partial\boldsymbol{\alpha}'}\right]^{-1}\boldsymbol{S}(\boldsymbol{\alpha}^{(t-1)}), \quad t = 1, 2, \ldots,$$

until $\boldsymbol{\alpha}^{(t)}$ converges to $\hat{\boldsymbol{\alpha}}$, say, where $\boldsymbol{\alpha}^{(t)}$ is the updated value of $\boldsymbol{\alpha}$ at the tth iteration.

Stage 2: Estimation of the covariance matrices $\boldsymbol{V}_i$

Let the $m \times m$ diagonal matrix $\tilde{\boldsymbol{W}}_i = \text{diag}(1, 1, \ldots, 1)$ be the working matrix, $i = 1, 2, \ldots, n$, and replace the covariance matrix $\boldsymbol{V}_i$ with the working matrix $\tilde{\boldsymbol{W}}_i$ in the estimating functions (5). For an initial value $\boldsymbol{\beta}^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)}, \ldots, \beta_p^{(0)})'$, apply the bisection method (Press et al., 1992, Chapter 9.1) to the function $U_{(k)}(\boldsymbol{\beta}_{(k)}^{(t)}, \hat{\boldsymbol{\alpha}})$ to obtain the tth update $\tilde{\beta}_k^{(t)}$ of β_k, $k = 1, 2, \ldots, p$; $t = 1, 2, \ldots$. Here the bisection method actually finds a zero crossing β_k (Jung, 1996). We cycle this process until $\tilde{\boldsymbol{\beta}}^{(t)} = (\tilde{\beta}_1^{(t)}, \tilde{\beta}_2^{(t)}, \ldots, \tilde{\beta}_p^{(t)})'$ converges to $\tilde{\boldsymbol{\beta}}$, say. Therefore, $\boldsymbol{V}_i$ in equation (5) may be estimated by the empirical estimate $\tilde{\boldsymbol{V}}_i$, where the (j, k) element is given by the empirical estimate

$$\tilde{p}_{ijk} = \frac{\sum_{i=1}^{n} I(R_{ij}=1)I(R_{ik}=1)I(Y_{ij} \geqslant \boldsymbol{x}_{ij}'\tilde{\boldsymbol{\beta}})I(Y_{ik} \geqslant \boldsymbol{x}_{ik}'\tilde{\boldsymbol{\beta}})}{\sum_{i=1}^{n} I(R_{ij}=1)I(R_{ik}=1)}, \quad j \neq k.$$

Stage 3: Estimation of parameters $\boldsymbol{\beta}$ of interest

Now we apply the estimating functions (5), with $\boldsymbol{V}_i$ replaced by $\tilde{\boldsymbol{V}}_i$ and $\boldsymbol{\alpha}$ replaced by $\hat{\boldsymbol{\alpha}}$, to update $\boldsymbol{\beta}$ using the bisection method in the same manner as in stage 2. Denote by $\hat{\boldsymbol{\beta}}$ the estimator of $\boldsymbol{\beta}$.

We comment that this estimation procedure is sensitive to the choice of initial value $\boldsymbol{\beta}^{(0)}$. The empirical estimate $\tilde{\boldsymbol{V}}_i$ may not be positive definite in some cases. A proper choice of initial values could avoid negative definite or singular matrix estimate $\tilde{\boldsymbol{V}}_i$.

3.3 *Asymptotic Properties*

Let $\mathcal{B}$ and $\mathcal{C}$ be the parameter spaces for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, respectively, and $\boldsymbol{A}_n(\boldsymbol{\beta},\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^{n}\phi^{-2}\cdot \boldsymbol{D}_i'\boldsymbol{\Gamma}_i\boldsymbol{V}_i^{-1}\boldsymbol{\Delta}_i\boldsymbol{\Gamma}_i\boldsymbol{D}_i$. Assume that $\boldsymbol{A}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \lim_{n\to\infty}\boldsymbol{A}_n(\boldsymbol{\beta},\boldsymbol{\alpha})$ is positive definite in a neighborhood of the true value $(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)$ of $(\boldsymbol{\beta},\boldsymbol{\alpha})$. Let $\boldsymbol{P} = -\phi\boldsymbol{A}(\boldsymbol{\beta},\boldsymbol{\alpha})$ and $\boldsymbol{Q}_i = \boldsymbol{U}_i - \mathrm{E}(\partial\boldsymbol{U}_i/\partial\boldsymbol{\alpha}')\cdot[\mathrm{E}(\partial\boldsymbol{S}_i(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}')]^{-1}\cdot\boldsymbol{S}_i(\boldsymbol{\alpha})$, both evaluated at $(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)$. In the Web Appendix we prove the following asymptotic properties.

THEOREM 1: *Subject to regularity conditions including* (R.1) *and* (R.3) *in Robins et al.* (1995) *and the assumptions in Jung* (1996), *we have the following results: as* $n\to\infty$,

(1) $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$
(2) $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{P}^{-1}\boldsymbol{\Sigma}[\boldsymbol{P}^{-1}]')$, *where* $\boldsymbol{\Sigma} = \mathrm{E}(\boldsymbol{Q}_i\boldsymbol{Q}_i')$.

Comparing to the asymptotic distribution of the estimator obtained from the mean regression model (e.g., Section 4 of Yi and Cook, 2002a), we may notice that the asymptotic distributions for the mean and median estimators differ in the form of the matrix $\boldsymbol{P}$. In the case of mean regression $\boldsymbol{P}$ equals the derivative matrix $\mathrm{E}(\partial\boldsymbol{U}_i/\partial\boldsymbol{\beta}')$; whereas for median regression this equality is no longer true because $E(\partial\boldsymbol{U}_i/\partial\boldsymbol{\beta}')$ does not exist due to discontinuity of $\boldsymbol{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})$. However, if we follow the spirit of Godambe and Thompson (1984) to define $E(\partial\boldsymbol{U}_i/\partial\boldsymbol{\beta}')$ for discontinuous estimating functions, we may obtain the following Theorem 2, which says that the asymptotic distributions for both the mean and median estimators assume the same presentation. The proof of Theorem 2 is given in the Web Appendix.

THEOREM 2: $\boldsymbol{P} = \mathrm{E}(\partial\boldsymbol{U}_i/\partial\boldsymbol{\beta}')$.

To conduct inference we need to estimate $\boldsymbol{P}$ and $\boldsymbol{\Sigma}$ as they contain the unknown parameters. We may estimate $\boldsymbol{P}$ by $-\hat{\phi}\boldsymbol{A}_n(\boldsymbol{\beta},\boldsymbol{\alpha})|_{(\boldsymbol{\beta},\boldsymbol{\alpha},\phi)=(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\hat{\phi})}$, where $\hat{\phi}$ is a consistent estimate of ϕ. Applying the law of large numbers element-wise, we obtain that $\mathrm{E}(\partial\boldsymbol{S}_i/\partial\boldsymbol{\alpha}')$ is consistently estimated by $\hat{\boldsymbol{S}}_\alpha = (\frac{1}{n}\sum_{i=1}^{n}(\partial\boldsymbol{S}_i/\partial\alpha_1),\ldots,\frac{1}{n}\sum_{i=1}^{n}(\partial\boldsymbol{S}_i/\partial\alpha_q))|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}}$, and $\mathrm{E}(\partial\boldsymbol{U}_i/\partial\boldsymbol{\alpha}')$ is estimated by $\hat{\boldsymbol{U}}_\alpha = (\frac{1}{n}\sum_{i=1}^{n}(\partial\boldsymbol{U}_i/\partial\alpha_1),\ldots,\frac{1}{n}\sum_{i=1}^{n}(\partial\boldsymbol{U}_i/\partial\alpha_q))|_{(\boldsymbol{\beta},\boldsymbol{\alpha},\phi)=(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\hat{\phi})}$, as $n\to\infty$, where $\partial\boldsymbol{U}_i/\partial\alpha_j = \phi^{-1}\boldsymbol{D}_i'\boldsymbol{\Gamma}_i\boldsymbol{V}_i^{-1}\cdot(\partial\boldsymbol{\Delta}_i/\partial\alpha_j)\cdot\boldsymbol{\epsilon}_i$, $\partial\boldsymbol{\Delta}_i/\partial\alpha_j = \mathrm{diag}(a_{ijk}, k=1,2,\ldots,m)$, and $a_{ijk} = -(I(R_{ik}=1)/\pi_{ik}^2)\cdot(\partial\pi_{ik}/\partial\alpha_j)$. As a result, $\boldsymbol{Q}_i$ is consistently estimated by, as $n\to\infty$, $\hat{\boldsymbol{Q}}_i = \hat{\boldsymbol{U}}_i - \hat{\boldsymbol{U}}_\alpha[\hat{\boldsymbol{S}}_\alpha']^{-1}\boldsymbol{S}_i(\hat{\boldsymbol{\alpha}})$ with $\hat{\boldsymbol{U}}_i = \hat{\phi}^{-1}\boldsymbol{D}_i'\boldsymbol{\Gamma}_i\boldsymbol{V}_i^{-1}\times\boldsymbol{\Delta}_i\boldsymbol{\epsilon}_i|_{(\boldsymbol{\beta},\boldsymbol{\alpha})=(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}})}$, and hence leading to a consistent estimate $\hat{\boldsymbol{\Sigma}} = n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{Q}}_i\cdot\hat{\boldsymbol{Q}}_i'$ of $\boldsymbol{\Sigma}$.

Finally, we comment that ϕ comes into play when forming the estimate of the covariance matrix. A consistent estimate of ϕ is given by

$$\hat{\phi}^{-1} = \left\{2h\sum_{i=1}^{n}\sum_{j=1}^{m}I(R_{ij}=1)\right\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}I(R_{ij}=1) \times (f(\hat{\mu}_{ij}))^{-1}\{I(Y_{ij}\geqslant\hat{\mu}_{ij}-h)-I(Y_{ij}\geqslant\hat{\mu}_{ij}+h)\},$$

where $\hat{\mu}_{ij}$ is the estimate of μ_{ij} with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}$, and h satisfies $h\to 0$ and $nh/\log(n)\to\infty$ as $n\to\infty$ (Jung, 1996).

4. Application to CD4+ Cell Counts Data

We apply the proposed method to the motivating data described in Section 1. There were 892 individuals in the study, in which 461 individuals were randomly assigned to the 1500-mg Zidovudine (treatment) group and 431 were placed in a placebo group (Volberding et al., 1990; Davis, 2002, p. 369). The CD4+ cell counts were measured, respectively, at weeks 8, 16, 32, and 48 after randomization. During the study a portion of participants was lost to follow-up. In Table 1, we report summary statistics for both treatment and control arms.

CD4+ cell counts are a biomarker for AIDS or AIDS-related complex disease. In this study, one primary objective is to assess the treatment effect on increasing CD4+ cell counts and possible time effect. As revealed in Figure 1 and Table 1, the outcomes are skewed with a number of outliers; therefore, we adopt a median regression model to postulate the measurement process. Treatment indicator, time factor, and their interaction term are included as covariates. Specifically, let Y_{ij} be the CD4+ cell counts for subject i at time j, $i=1,2,\ldots,892$, $j=1,2,3,4$, then the median regression model is taken as

$$\mu_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2j} + \beta_3 x_{i1}x_{i2j}, \tag{6}$$

where $x_{i1}=1$ if subject i received treatment, and 0 otherwise; and $x_{i2j}=j$ indexes the jth assessment time for subject i.

As suggested in Volberding et al. (1990), dropouts may reflect the selective withdrawal from the study of subjects with low or declining CD4+ cell counts. The higher number of withdrawals from the placebo group was reportedly due

Table 1
Summary statistics for the CD4+ cell count data

	Time point	1	2	3	4
Treatment	Mean	403.2	407.8	407.6	390.5
	Median	390.0	388.5	394.0	378.0
	Missingness	0%	5.9%	15.8%	36.9%
Placebo	Mean	373.4	365.8	380.6	401.0
	Median	360.0	345.0	357.0	384.0
	Missingness	0%	7.0%	21.6%	42.2%

Table 2
Analyses of the CD4+ cell count data

Parameter	Est.	SEs	95% C.I.	p-value
Response				
Intercept (β_0)	343.001	5.037	(333.128, 352.873)	<0.001
Treatment (β_1)	56.999	7.059	(43.164, 70.834)	<0.001
Time (β_2)	3.500	1.707	(0.154, 6.846)	0.040
Interaction (β_3)	−10.500	2.435	(−15.273, −5.726)	<0.001
Missingness				
Intercept (α_0)	1.215	0.148	(0.924, 1.506)	<0.001
Past resp. (α_1)	0.0013	0.0004	(0.0006, 0.0021)	<0.001
Treatment (α_2)	0.145	0.107	(−0.064, 0.355)	0.174

to a possible desire to take treatment to increase CD4+ cell counts. To feature the dependence on the observed outcomes and the treatment status, here we postulate the dropout process with the logistic regression model

$$\text{logit}\lambda_{ij} = \alpha_0 + \alpha_1 y_{i,j-1} + \alpha_2 x_{i1}. \quad (7)$$

Table 2 displays the estimates, standard errors (SEs), 95% confidence intervals, and p-values for the model parameters. The estimates of α_j's are displayed at the bottom of Table 2. As can be seen, α_1 is statistically significant with a positive estimate. If CD4+ cell counts are observed increasing at the previous assessment, individuals are more likely to be present for the following assessment. Positive estimate of α_2 indicates that a subject is more likely to drop out of the study if he/she did not receive a treatment, but this trend is not statistically significant. The top panel of Table 2 reports on the results for the covariate effects for the measurement process. The analysis reveals a strongly positive treatment effect on increasing CD4+ cell counts. There is no ample evidence of time effect on increasing CD4+ cell counts. However, there does exists evidence to support the interaction effect of time and treatment on increasing CD4+ cell counts.

Finally, we comment that we should be cautious in interpreting the analysis results. Missing data mechanisms are generally not testable unless certain restrictions are made. Postulating the dropout process by a model like (7) may help us understand the impact of dropouts on estimation of $\boldsymbol{\beta}$ parameters.

5. Simulation Study

5.1 *Assessment of the Proposed Median Regression Method*

Lipsitz et al. (1997) discussed a special case of estimating functions (5), where the covariance matrix $\boldsymbol{V}_i$ is simply taken as the identity matrix $\boldsymbol{I}_{m\times m}$. That form, though still being unbiased for a given $\boldsymbol{\alpha}$, failed to account for serial correlation among response components. Moreover, theoretical properties of the resulting estimators have not been established in Lipsitz et al. (1997). In this subsection, we evaluate the performance of the proposed method through simulation studies, in contrast to the method discussed in Lipsitz et al. (1997). Here analysis 1 refers to the proposed method whereas analysis 2 is based on the method discussed in Lipsitz et al. (1997).

We simulate longitudinal responses from a multivariate normal distribution with marginal medians (equation (6)) and the covariance matrix $\sigma^2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the $m \times m$ matrix with the diagonal elements being 1 and the off diagonal elements ρ. The covariates include a treatment indicator x_{i1}, simulated from Bernoulli(0.5), the temporal effect x_{i2j}, which is the index j of the follow-up time, and the interaction term $x_{i1}x_{i2j}$. We consider a number of scenarios with $\rho = 0$, 0.3, 0.5, and 0.7 to feature increasing strength of association from independence to high correlation. σ^2 is taken as 1.0. Set $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)' = (6.0, -5.0, 1.0, 15.0)'$ and $m = 6$. Take $n =$ 200 and 1000 to represent moderate and large sample sizes, respectively. Five hundred simulations are conducted for each parameter configuration. For the dropout process we employ logistic regression model (7), where parameters $\boldsymbol{\alpha}$ are taken as $(1.0, 0.1, -0.5)'$.

In Table 3, we report the differences between the estimates and the true values along with the empirical SEs and the empirical coverage rates for 95% confidence intervals. As expected, the differences between the estimates and the true

Table 3
Assessment of the proposed median regression method

			Estimate − true value (SEs)				Coverage percentage			
Size	Analysis	ρ	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
200	Analysis 1	0.0	0.072 (0.106)	−0.065 (0.145)	−0.014 (0.031)	0.019 (0.040)	94.2	96.0	94.0	94.0
		0.3	0.041 (0.120)	−0.031 (0.166)	−0.005 (0.032)	0.008 (0.042)	94.5	96.1	95.6	96.1
		0.5	0.037 (0.125)	−0.023 (0.174)	−0.004 (0.033)	0.005 (0.041)	93.8	95.9	94.0	96.4
		0.7	0.043 (0.128)	−0.023 (0.174)	−0.001 (0.034)	0.002 (0.040)	94.8	96.4	94.2	93.2
	Analysis 2	0.0	0.058 (0.100)	−0.081 (0.134)	−0.015 (0.029)	0.022 (0.038)	95.4	95.8	95.2	95.0
		0.3	0.046 (0.108)	−0.075 (0.147)	−0.011 (0.029)	0.019 (0.037)	95.2	95.4	96.0	97.0
		0.5	0.048 (0.109)	−0.065 (0.147)	−0.012 (0.029)	0.017 (0.035)	96.2	97.0	96.0	97.8
		0.7	0.055 (0.113)	−0.078 (0.154)	−0.011 (0.028)	0.017 (0.035)	96.2	96.8	96.0	96.4
1000	Analysis 1	0.0	0.026 (0.058)	−0.025 (0.080)	−0.005 (0.016)	0.006 (0.022)	93.8	96.8	96.4	95.8
		0.3	0.012 (0.062)	−0.010 (0.082)	−0.001 (0.016)	0.001 (0.021)	95.2	96.4	95.4	95.6
		0.5	0.010 (0.056)	−0.002 (0.077)	−0.002 (0.015)	0.002 (0.018)	95.8	96.8	95.3	95.5
		0.7	0.005 (0.060)	−0.004 (0.082)	0.001 (0.014)	−0.001 (0.018)	96.7	95.3	95.3	94.7
	Analysis 2	0.0	0.025 (0.057)	−0.030 (0.079)	−0.006 (0.016)	0.007 (0.022)	94.6	96.4	96.2	95.2
		0.3	0.026 (0.058)	−0.025 (0.080)	−0.005 (0.016)	0.006 (0.022)	93.2	96.4	96.0	95.6
		0.5	0.025 (0.058)	−0.026 (0.075)	−0.007 (0.015)	0.008 (0.018)	93.2	95.6	93.8	93.8
		0.7	0.020 (0.060)	−0.025 (0.081)	−0.004 (0.014)	0.005 (0.018)	94.6	94.0	94.0	94.0

values and SEs obtained from both analyses become smaller as sample size increases. As $\rho = 0$, both analyses give rise to comparable differences between the estimates and the true values. However, when the association exists among the response components, the differences between the estimates and the true values yielded by the two analyses could be quite different. Analysis 1 tends to result in a lot smaller differences than analysis 2, and the magnitude in the two analyses is more noticeable as the association among the response components becomes stronger. It is not surprising that the differences between the estimates and the true values resulted from analysis 2 do not appear to change much with the strength of ρ, because the association among the response components is not accounted for by analysis 2. Finally, we notice that the empirical coverage rates are reasonably consistent with the nominal level 95% in general.

5.2 *Comparison of Median and Mean Regression Methods*

In practice, the distribution of data could be heavily tailed or highly skewed. In this subsection, we assess the performance of the proposed method under a variety of distributions. Typically, we undertake simulations to compare the performance of the proposed median regression method in contrast to the usual mean regression approach.

In addition to a normal distribution, we consider nonnormal distributions—Cauchy and exponential distributions, which represent distinct scenarios of heavily tailed or skewed distributions. We simulate data from each of the distributions with the median regression given by equation (6). To be specific, when considering an exponential distribution we generate Y_{ij} from the density function $f(y) = (1/\theta_{ij})\exp(-y/\theta_{ij})$ with $\theta_{ij} = (\log(2))^{-1}\mu_{ij}$; when generating data from a Cauchy distribution we invoke the cumulative distribution function $F(y) = 1/2 + (1/\pi)\arctan(y - \mu_{ij})$ for Y_{ij}; and we employ the distribution $N(\mu_{ij}, 1)$ to simulate Y_{ij} for the case of a normal distribution. Again, m is set as 6, and the covariates are simulated as in Section 5.1. The vector of covariate coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ is set to be $(0.5, 0.5, 0.1, 0.1)'$. Missing data indicators are generated from logistic regression model (7), where parameters $\boldsymbol{\alpha}$ are taken as $(2.0, 0.5, 0.5)'$. Two hundred simulations are run for each of the distributions for the cases with $n = 200$ and $n = 1000$.

We fit the data with the proposed median regression method as well as the IPWGEE approach based on mean regression. In Table 4, we report the differences between the estimates and the true values, the empirical SEs for $\hat{\boldsymbol{\beta}}$, and the coverage rates for 95% confidence intervals. For the normal distribution both methods provide fairly comparable finite sample differences between the estimates and the true values, though the median regression method produces relatively larger differences for some parameters. Median regression is less efficient than mean regression as it results in larger SEs. But median regression does yield reasonable coverage rates for 95% confidence intervals just like mean regression does. As the sample size increases, the SEs obtained from both methods reduce as expected.

The performance for the median and mean regression methods, however, is dramatically different for nonnormally distributed data. If data are generated from a Cauchy distribution, mean regression breaks down as the moments for Cauchy distributions do not exist. Nevertheless, median regression performs reasonably well. The finite sample differences between the estimates and the true values are fairly small, and the coverage rates for 95% confidence intervals are satisfactorily close to the nominal level. Again, it is observed that a larger sample size leads to smaller SEs. For data generated from an exponential distribution, median regression still maintains reasonable performance with relatively satisfactory coverage rates and finite sample differences between the estimates and the true values. However, mean regression fails to provide comparable results. The differences between the estimates and the true values are consistently and considerably larger than those from median regression, and the coverage rates are far from the nominal level. It is seen that mean regression may never capture the true values for some parameters.

In summary, the proposed median regression method performs satisfactorily for a variety of data with distinct distributions. The finite sample differences between the estimates and the true values are reasonably small, and the coverage rates are relatively reliable. The performance of the weighted mean regression approach, however, does remarkably depend on the shape of the distribution of the data. For normally distributed data, mean regression produces reasonably

Table 4
Comparison of median and mean regression methods under different distributions

			Estimate − true value (SEs)				Coverage percentage			
Size	Distribution	Method	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
200	Normal	Median	0.010 (0.211)	−0.030 (0.291)	−0.001 (0.057)	0.001 (0.078)	94.0	96.5	95.5	96.0
		Mean	0.012 (0.098)	−0.001 (0.137)	−0.002 (0.027)	−0.002 (0.037)	96.0	96.0	96.5	96.5
	Exponential	Median	0.029 (0.670)	−0.163 (0.893)	−0.001 (0.131)	0.065 (0.166)	99.4	98.9	98.3	92.5
		Mean	0.222 (0.076)	0.226 (0.164)	0.044 (0.022)	0.044 (0.042)	12.5	72.5	52.5	86.0
	Cauchy	Median	0.087 (0.242)	0.019 (0.412)	0.012 (0.080)	0.057 (0.157)	94.6	96.7	92.4	95.7
		Mean	–	–	–	–	–	–	–	–
1000	Normal	Median	0.017 (0.112)	−0.028 (0.155)	−0.004 (0.031)	0.007 (0.042)	95.5	93.5	93.5	95.5
		Mean	−0.001 (0.045)	0.005 (0.062)	0.001 (0.012)	−0.002 (0.017)	94.0	95.0	95.0	94.0
	Exponential	Median	0.000 (0.210)	−0.018 (0.291)	0.011 (0.032)	0.005 (0.042)	99.5	97.9	97.3	92.5
		Mean	0.225 (0.035)	0.213 (0.074)	0.045 (0.010)	0.043 (0.019)	0.0	17.5	0.0	30.5
	Cauchy	Median	0.039 (0.194)	0.045 (0.284)	0.009 (0.065)	0.014 (0.082)	96.3	96.3	96.3	96.3
		Mean	–	–	–	–	–	–	–	–

accurate results, but it may break down or fail to provide reliable results for data with a distribution far from a normal one.

6. Extension to Estimation of Association Parameters

In this section, we discuss an extension for which the association structure of the repeated measurements is specifically incorporated in estimation procedures. This development is useful when the interest also lies in characterizing association among response components (e.g., Prentice, 1988; Yi and Cook, 2002a; Yi and Thompson, 2005).

For $i = 1, 2, \ldots, n$, let $Z_{ijk} = I(Y_{ij} \geqslant \mu_{ij})\ I(Y_{ik} \geqslant \mu_{ik})$ for $j < k$, and $\boldsymbol{Z}_i = (Z_{i12},\ Z_{i13}, \ldots, Z_{i1m},\ Z_{i23}, \ldots, Z_{i2m}, \ldots, Z_{i,m-1,m})'$. Define the odds ratio

$$\psi_{ijk} = \frac{P(Y_{ij} \geqslant \mu_{ij}, Y_{ik} \geqslant \mu_{ik} \,|\, \boldsymbol{x}_i) P(Y_{ij} < \mu_{ij}, Y_{ik} < \mu_{ik} \,|\, \boldsymbol{x}_i)}{P(Y_{ij} \geqslant \mu_{ij}, Y_{ik} < \mu_{ik} \,|\, \boldsymbol{x}_i) P(Y_{ij} < \mu_{ij}, Y_{ik} \geqslant \mu_{ik} \,|\, \boldsymbol{x}_i)}, \quad j < k,$$

which can facilitate the association between Y_{ij} and Y_{ik}, in combination of the covariate information. Regression models for the association can be typically specified as

$$\log \psi_{ijk} = \boldsymbol{u}_{ijk}^{*\prime} \boldsymbol{\delta}, \tag{8}$$

where $\boldsymbol{u}_{ijk}^{*}$ is a vector of covariates, which specifies the form of the association between Y_{ij} and Y_{ik}, and $\boldsymbol{\delta}$ is a vector of regression parameters. Letting $\boldsymbol{u}_{ijk}^{*}$ be the scalar one, for example, leads to the exchangeable association between responses within the same subject; and setting $\boldsymbol{u}_{ijk}^{*} = |j - k|$ gives rise to an autoregressive association.

The joint probability p_{ijk} can be written as:

$$p_{ijk} = \begin{cases} \dfrac{a_{ijk} - \left[a_{ijk}^2 - 4\psi_{ijk}(\psi_{ijk} - 1)\mu_{ij}\mu_{ik}\right]^{\frac{1}{2}}}{2(\psi_{ijk} - 1)}, & \text{if } \psi_{ijk} \neq 1, \\ \mu_{ij}\mu_{ik}, & \text{if } \psi_{ijk} = 1, \end{cases} \tag{9}$$

where $a_{ijk} = 1 - (1 - \psi_{ijk})(\mu_{ij} + \mu_{ik})$. Let $\boldsymbol{p}_i = (p_{i12}, p_{i13}, \ldots, p_{i1m},\ p_{i23}, \ldots, p_{i2m}, \ldots, p_{i,m-1,m})'$, $\boldsymbol{C}_i = \partial \boldsymbol{p}_i / \partial \boldsymbol{\delta}'$, $\boldsymbol{W}_i^{*} = \mathrm{diag}(p_{ijk}(1 - p_{ijk}),\ j < k)$, and $\boldsymbol{\Delta}_i^{*} = \mathrm{diag}(I(R_{ij} = 1)I(R_{ik} = 1)/\pi_{ijk},\ j < k)$, where $\pi_{ijk} = P(R_{ij} = 1, R_{ik} = 1 \,|\, x_i, y_i)$. If $\boldsymbol{U}_i^{*} = \boldsymbol{C}_i' [\boldsymbol{W}_i^{*}]^{-1} \boldsymbol{\Delta}_i^{*} (\boldsymbol{Z}_i - \boldsymbol{p}_i)$, then association parameters $\boldsymbol{\delta}$ can be estimated by the estimating equations:

$$\boldsymbol{U}^{*} = \sum_{i=1}^{n} \boldsymbol{U}_i^{*} = \boldsymbol{0}.$$

It is easily seen that $E(\boldsymbol{U}^{*}) = \boldsymbol{0}$, i.e., $\boldsymbol{U}^{*}$ consists of unbiased estimating functions for $\boldsymbol{\delta}$. Estimation of both $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ can be performed based on $\boldsymbol{U}$ and $\boldsymbol{U}^{*}$ by adapting the algorithm described in Section 3.2, where π_{ijk} is replaced by its estimate $\hat{\pi}_{ijk}$ that can be handled by adapting the argument in Yi and Cook (2002a).

7. Discussion

This article discusses median regression models for longitudinal data with dropouts. Inverse probability weighted estimating equations are employed to estimate median regression parameters. This perhaps appears similar to the traditional approach of the inverse probability weighted estimating equations for mean regression models that have been extensively discussed in the literature. However, our current development cannot be viewed as a direct application of that approach, because the constructed estimating functions here are not continuous as usual. In this article, we have established asymptotic properties for the estimators obtained from the proposed weighted estimating equations. The simulation studies demonstrate that the proposed method performs well under a variety of situations, especially when mean regression breaks down for distributions that are far from normal distributions.

The proposed methods can be extended in a straightforward fashion to accommodate general quantile regression models, which may be of practical interest sometimes. He et al. (2003) considered the LAD approach for median regression models for complete longitudinal data. It is interesting to adapt the proposed methods to the LAD regressions to handle incomplete longitudinal data.

8. Supplementary Materials

A Web Appendix referenced in Section 3.3 is available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

The authors would like to thank Professor V. P. Godambe for his helpful conversations that inspired this work. The authors gratefully acknowledge the editor, two anonymous referees, and especially an associate editor for their helpful comments, which greatly improved the form of this manuscript. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

Bang, H. and Tsiatis, A. (2002). Median regression with censored cost data. *Biometrics* **58,** 643–649.

Cook, R. J., Zeng, L., and Yi, G. Y. (2004). Marginal analysis of incomplete longitudinal binary data: A cautionary note on LOCF imputation. *Biometrics* **60,** 820–828.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. New York: Springer-Verlag, Inc.

Fitzmaurice, G. M., Molenberghs, G., and Lipsitz, S. R. (1995). Regression models for longitudinal binary responses with informative drop-outs. *Journal of the Royal Statistical Society, Series B* **57,** 691–704.

Godambe, V. P. (2001). *Estimation of median: Quasi-likelihood and optimum estimating functions*. Working Paper 01, University of Waterloo, Waterloo, Ontario.

Godambe, V. P. and Thompson, M. E. (1984). Robust estimation through estimating equations. *Biometrika* **71,** 115–125.

He, X., Fu, B., and Fung, W. K. (2003). Median regression for longitudinal data. *Statistics in Medicine* **22,** 3655–3669.

Hogan, J. W. and Lee, J. Y. (2004). Marginal structure quantile models for longitudinal observational studies with time-varying treatment. *Statistica Sinica* **14,** 927–944.

Jung, S. H. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91,** 251–257.

Kenward, M. G. (1998). Selection models for repeated measurements with nonrandom dropout: An illustration of sensitivity. *Statistics in Medicine* **7,** 2723–2732.

Koenker, R. and Bassett, G. W. (1978). Regression quantile. *Econometrica* **46,** 33–50.

Koenker, R. and Xiao, Z. (2001). Inference on the quantile regression process. *Econometrica* **70,** 1583–1612.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G., and Zhao, L. P. (1997). Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Applied Statistics* **46,** 463–476.

Morgenthaler, S. (1992). Least-absolute-deviations fits for generalized linear models. *Biometrika* **79,** 747–754.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, D. McFadden and R. Engler (eds), Volume 4. Amsterdam: North-Holland.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association* **98,** 1001–1012.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44,** 1033–1048.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*, 2nd edition. New York: Cambridge University Press.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90,** 106–121.

Volberding, P. A., Lagakos, S. W., Koch, M. A., Pettinelli, C., Myers, M. W., Booth, D. K., Balfour, H. H., Reichman, R. C., Bartlett, J. A., Hirsch, M. S., Murphy, R. L., Hardy, W. D., Soeiro, R., Fischl, M. A., Bartlett, J. G., Merigan, T. C., Hyslop, N. E., Richman, D. D., Valentine, F. T., and Gorey, L. (1990). Zidovudine in asymptomatic human immunodeficiency virus infection: A controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter. *The New England Journal of Medicine* **322,** 941–949.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association* **94,** 137–145.

Yi, G. Y. and Cook, R. J. (2002a). Marginal methods for incomplete longitudinal data arising in clusters. *Journal of the American Statistical Association* **97,** 1071–1080.

Yi, G. Y. and Cook, R. J. (2002b). Second order estimating equations for clustered longitudinal binary data with missing observations. In *Recent Advances in Statistical Methods*, Y. P. Chaubey (ed), 352–366. London: World Scientific Publishing Company, Inc.

Yi, G. Y. and Thompson, M. E. (2005). Marginal and association regression models for longitudinal binary data with drop-outs: A likelihood-based approach. *The Canadian Journal of Statistics* **33,** 3–20.

Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90,** 178–184.

Received May 2005. *Revised May* 2008.
Accepted May 2008.