Robust inference for longitudinal data analysis with non-ignorable and non-monotonic missing values

Chi-hong Tseng*, Robert Elashoff, Ning Li and Gang Li

A common problem in the longitudinal data analysis is the missing values due to subject's missed visits and loss to follow up. Although many novel statistical approaches have been developed to handle such data structures in recent years, few methods are available to provide robust inference in the presence of outlying observations. In this paper we propose two methods, t-distribution model and robust normal model, for robust inference with non-ignorable non-monotonic missing data problems in longitudinal studies. These methods are conceptually simple and computationally straight forward. We also conduct simulation studies and use a real data example to demonstrate the performance of these methods.

Keywords and phrases: Pseudo-likelihood, t-distribution, Huber function, Missing not at random.

1. INTRODUCTION

Longitudinal studies are frequently used in clinical research to monitor disease progression and treatment effects over time. The most commonly used model to analyse continuous longitudinal data is the linear mixed-effects model proposed by Laird et al. (1982). This model specifies flexible random effects to describe the within-subject correlation present in the longitudinal data. However, the normality assumption in the linear mixed effects model is often unrealistic and its performance may be compromised when the underlying normality assumption is violated (Pinheiro et al., 2001).

To relax the restricted normality assumption in longitudinal data analysis, various robust approaches have been proposed in the context of likelihood inference. Huggins (1993) and Richardson et al. (1995) considered reweighting each observation's contribution to the likelihood function and score equations to reduce the impact of outliers. Similarly, Gill (2000) used the robust version of the likelihood function and discussed its inference on regression coefficients, covariance parameters estimation, prediction and

*Corresponding author.

model selection. Others used a broader family of distributions with less restriction to allow flexible modelling of random effects and measurement error distributions. For example Pinheiro et al. (2001) proposed linear mixed effects models with a t distribution for random effects and measurement errors. Verbeke et al. (1996) used a mixture of normal distributions to model random effects. They also proposed statistical tests for heterogeneity in the random-effects distribution. Zhang et al. (2001) considered the semi-nonparametric (SNP) method that allows the random effects to follow a density belonging to a class of smooth densities. Ghidey et al. (2004) proposed the penalized Gaussian mixture model for random effects.

Another concern in longitudinal data analysis is the missing data problem. During the course of study follow-up, the outcome variables and covariates can be missing due to subjects' non-response, dropout, death or other reasons. There is a rich statistical literature on the analysis of missing data. Little et al. (2002) defined and discussed various missing data mechanisms. If the missingness is independent of the observed and unobserved data, the missing mechanism is missing completely at random (MCAR). If given the observed data, the missingness is independent of unobserved data, it is defined as missing at random (MAR). With MCAR or MAR data, the missing data mechanism (the distribution of the missing data) can be ignored when using likelihood methods, and the missing data mechanism is considered ignorable (Little et al., 2002). However, in practice it may be difficult to justify the independence assumptions of MCAR or MAR, and we need to consider missing not at random (MNAR), where the missing probability depends on unobserved data. For example, in our motivating example of the Scleroderma Lung Study, about 60% of patient dropouts are due to death or treatment failure, which are likely related to lack of treatment efficacy.

With non-ignorable missing data, statistical modelling of the missing mechanisms along with the modelling of the outcome of interest is generally required for unbiased inference (Hogan et al., 1981; Ibrahim et al., 2009). Three modelling strategies are commonly used to handle non-ignorable missing data: selection models, pattern mixture models and shared parameter models. Selection models (Wu et al., 1988; Schluchter, 1992) specify an overall outcome model and

a missing mechanism model that describes how missing data depend on the outcome variable. They provide a natural way to express the outcome process and the missing mechanism. Pattern mixture models (Little, 1993) consider the full data as a mixture of data from different missing data patterns. They allow different outcome models for subjects with different missing data patterns. Finally, shared parameter models use latent variables, such as random effects, to account for the dependence between the outcome variable and missingness. For example, Elashoff et al. (2007, 2008) used a joint modelling approach to analyse the lung function data in a Scleroderma study in the presence of non-ignorable dropouts. They used a cause-specific hazard frailty model for competing risk failure times, and a linear mixed effects model for the outcome variable. The missing data caused by the non-ignorable dropouts in the linear mixed effects model are adjusted by the correlated random effects from the frailty model and linear mixed effects model.

Two types of missing data patterns are generally distinguished (Ibrahim et al., 2009) in the statistical literature. The first type is called "intermittently missing" or "not monotonic missing", where a subject may miss particular visits during the course of study follow-up and return at later scheduled visits. On the other hand, "dropout" or "monotonic missing" data refer to a subject who leaves the study at some point and never returns. Few methods are available for non-monotonic and non-ignorable missing data. Ibrahim et al. (2001) used a full likelihood approach that specifies the joint likelihood of the outcome and missing indicator to handle non-monotonic and non-ignorable missingness. The full likelihood approach requires a full specification of the joint distribution of the data and missingness. It is also computationally intensive, although efficient if the specified model is true. Troxel et al. (1998) and Parzen et al. (2006) proposed the pseudo-likelihood approach for the non-monotonic and non-ignorable missing data problem. It requires fewer modelling assumptions and simplifies computational effort.

1.1 Relationship to current literature

In this paper we are interested in making robust statistical inference that accommodates outlying observations with non-monotonic and non-ignorable missing data in longitudinal studies with continuous outcome variables. A number of likelihood based extensions to incorporate robust inference have been proposed for the non-ignorable missing data problem. For example, Brown et al. (2003) relaxed the distributional assumption of the random effects in the longitudinal model. Li et al. (2009) used a t-distribution to model the distribution of measurement and model errors; this model effectively down-weights the influence of outlying observations. However, these methods are computationally intensive and not applicable to non-monotonic missing data.

We extend the pseudo-likelihood approach (Troxel et al., 1998; Parzen et al., 2006) with a t-distribution model

(Lange et al., 1989; Pinheiro et al., 2001; Li et al., 2009) and a robust normal model (Huggins, 1993; Richardson et al., 1995; Gill, 2000) to make robust inference with non-monotonic and non-ignorable missing data. Troxel et al. (1998) used the pseudo-likelihood method to analyse continuous longitudinal data, assuming a normal distribution and allowing non-monotonic non-ignorable missing data. However, as we will demonstrate in our simulation study, the normal distribution assumption in Troxel et al. (1998) lacks robust properties such that the estimation is sensitive to outlying observations. Lange et al. (1989) and Pinheiro et al. (2001) discussed the use of t-distribution for robust inference without missing data. Gill (2000) used the robust normal model with full likelihood for robust inference without missing data. Li et al. (2009) modelled longitudinal and survival data simultaneously to handle monotonic missing data. They used the t distribution in the longitudinal model for robust inference.

Our pseudo-likelihood approach uses the selection model to handle the missing data problem; it consists of a linear model that specifies the marginal distribution of the outcome and a logistic regression model as the missing mechanism model that describes how missing data depend on the outcome variable. The pseudo-likelihood approach has several advantages. First, it only requires specification of the marginal distribution of the outcome variable, without making assumptions on the joint distribution of the data in the longitudinal model. Second, because the model specifies the marginal distribution of the outcome variable, it readily provides the population average interpretation of the estimates. Third, when compared to a full likelihood model, the pseudo-likelihood approach has a simpler formulation and is computationally straightforward. Our general model stated above is also mathematically tractable and computationally straightforward in comparison with other robust full likelihood methods (Huggins, 1993; Gill, 2000).

The paper is organized as follows: in section 2, we present the model specification and its inference based on the pseudo-likelihood approach; in section 3, we conduct simulation studies to examine the effectiveness and the robust properties of our methods; in section 4, we demonstrate the methods with a real example from a scleroderma clinical trial. We state the conclusions in section 5.

2. PSEUDO-LIKELIHOOD APPROACH

In this section, we formulate the selection model for non-ignorable missing data and the inference based on pseudo-likelihood approach according to Troxel et al. (1998) and Parzen et al. (2006). Let $Y_i(t)$ be the outcome variable of the ith subject at time t, and $R_i(t)$ be the response indicator of the ith subject at time t, for $i = 1, \ldots, n$, and $t = 1, \ldots, K$. $R_i(t) = 1$ if $Y_i(t)$ is observed, and $R_i(t) = 0$ if $Y_i(t)$ is missing. $X_i(t)$ and $W_i(t)$ are the predictive covariates

for $Y_i(t)$ and $R_i(t)$ respectively; they may be overlapped or disjoint. We assume that the outcome variable $Y_i(t)$ is subject to non-ignorable missing data and $X_i(t)$ and $W_i(t)$ are completely observed. To handle the non-ignorable missingness, the selection model specifies an overall outcome model and a missing mechanism model. The outcome variable $Y_i(t)$ follows the marginal linear model with predictive covariate $X_i(t)$:

$$(1) \qquad Y_i(t) = X_i(t)\beta + \sigma\epsilon_i(t).$$

We assume the marginal distribution of $Y_i(t)$ has a density function $f(Y_i(t)|X_i(t))$. The regression coefficients β are the parameters of interest and σ is a scale parameter. $\epsilon_i(t)$ represent the standardized residual and are correlated. For the missing mechanism model, we assume that $R_i(t)$ is associated with covariates $W_i(t)$ and current outcome $Y_i(t)$ with a logistic model.

$$(2) \quad \mathrm{Prob}(R_i(t) = 1) = \frac{\exp(\alpha_0 + W_i(t)\alpha_w + Y_i(t)\alpha_y)}{1 + \exp(\alpha_0 + W_i(t)\alpha_w + Y_i(t)\alpha_y)}$$

Equations (1) and (2) constitute our model.

The pseudo-likelihood (Gong et al., 1981) is then formulated by fixing the nuisance parameters of pairwise correlations at zero; the data of different time points are treated as independent. Let $\Theta = (\beta, \sigma, \alpha)$, the pseudo-likelihood and log pseudo-likelihood are given by,

$$L(\Theta)$$
$$= \prod_{i=1}^{n}\prod_{t=1}^{K}[f(Y_i(t)|X_i(t))f(R_i(t)|W_i(t),Y_i(t))]^{R_i(t)}$$
$$\times \left[\int f(Y_i(t)|X_i(t))f(R_i(t)|W_i(t),Y_i(t))dY_i(t)\right]^{1-R_i(t)},$$

$$\ell(\Theta) = \log L(\Theta) = \sum_{i=1}^{n}\ell_i(\beta,\sigma,\alpha) = \sum_{i=1}^{n}\sum_{t=1}^{K}\ell_i^t(\Theta),$$

and

$$\ell_i^t(\Theta) = R_i(t)[\log f(Y_i(t)|X_i(t)) \\ + \log f(R_i(t)|W_i(t),Y_i(t))] \\ + (1-R_i(t))\log\int f(Y_i(t)|X_i(t)) \\ \times f(R_i(t)|W_i(t),Y_i(t))dY_i(t)$$

The maximum pseudo-likelihood estimator $\hat{\Theta} = (\hat{\beta}, \hat{\sigma}, \hat{\alpha})$ maximizes the log pseudo-likelihood, and it can be obtained by solving the pseudo score equations:

$$(3) \qquad S(\Theta) = \frac{\partial}{\partial\Theta}\ell(\Theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\Theta}\ell_i(\Theta) = \sum_{i=1}^{n}S_i(\Theta) = 0$$

Troxel et al. (1998) demonstrated the consistency of the maximum pseudo likelihood estimator if the longitudinal model and missing model are correctly specified. We also provide an analysis of the consistency of our t distribution model in the Appendix. Because the observations from the same individual are likely correlated, we need to use the "sandwich" variance estimator (White, 1982; Liang et al., 1986).

$$n^{1/2}(\hat{\Theta} - \Theta) \longrightarrow N(0, \Sigma(\Theta))$$

where

$$\Sigma(\Theta)$$
$$= \left[\frac{1}{n}E\left(\frac{\partial(\Theta)}{\partial S(\Theta)}\right)\right]^{-1}\frac{1}{n}\sum E(S_i(\Theta)S_i'(\Theta))\left[\frac{1}{n}E\left(\frac{\partial S(\Theta)}{\partial(\Theta)}\right)\right]^{-1}$$

A consistent variance estimator can be obtained by $\hat{\Sigma} = \Sigma(\hat{\Theta})$.

The pseudo-likelihood approach in the missing data problem is also related to the expected score method proposed by Wang et al. (2001, 2008). For an observed $Y_i(t)$, its contribution to the score equation (3) is

$$(4)$$
$$S_i(t, \Theta) = \frac{\partial}{\partial\Theta}\ell_i^t(\Theta)$$
$$= \frac{\partial}{\partial\Theta}\log f(Y_i(t)|X_i(t)) + \log f(R_i(t)|W_i(t),Y_i(t)).$$

When $Y_i(t)$ is unobserved, its contribution to the score equation (3) has a nice mathematical expression as the conditional expectation of $S_i(t, \Theta)$ given the observed data $O_i(t) = (R_i(t) = 0, X_i(t), W_i(t))$.

$$(5) \quad \frac{\partial}{\partial\Theta}\int_{Y_i(t)} f(Y_i(t)|X_i(t))f(R_i(t) = 0|W_i(t),Y_i(t))dY_i(t)$$
$$= E\left[S_i(t, \Theta)|O_i(t)\right]$$

The overall pseudo score equations (3) combine the likelihood score (4) from the complete data, and the conditional expectation of score (5) for incomplete data. Furthermore, for the incomplete data, the derivative of the score can be expressed as a combination of conditional expectation of $S_i(t, \Theta)$ and its derivative $\frac{\partial}{\partial\Theta}S_i(t, \Theta)$:

$$(6) \qquad E\left[\frac{\partial}{\partial\Theta}S_i(t, \Theta)|O_i(t)\right] + E\left[S_i(t, \Theta)S_i'(t, \Theta)|O_i(t)\right]$$
$$- E\left[S_i(t, \Theta)|O_i(t)\right]E\left[S_i'(t, \Theta)|O_i(t)\right]$$

2.1 Robust inference

The first robust inference approach uses a t-distribution to characterize the marginal distribution of the outcome. The marginal distribution of $\epsilon_i(t)$ in equation (1) follows a t-distribution with v degrees of freedom; $\epsilon_i(t) \sim t-distribution(v)$, $t = 1, \ldots, K$. There are various ways to incorporate

the t distribution model for robust inferences (Lange et al., 1989). For example, given sufficient data, one can estimate v based on likelihood method. In this article, we use a fixed $v = 3$ throughout for robust inference. The t distribution with 3 degrees of freedom has sufficient long tails and provides a considerable degree of down-weighting for extreme outliers (Lange et al., 1989). It also simplifies the model and computational effort, as recommended by Tukey (1949) and Lange et al. (1989).

The second robust inference approach incorporates the robust normal likelihood (Huggins, 1993; Gill, 2000). The log-likelihood for an observed $Y_i(t)$ in the outcome model equation (1) under the normal distribution assumption is,

$$\log f(Y_i(t)|X_i(t))$$
$$= \text{constant} - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\left(\frac{Y_i(t) - X_i(t)\beta}{\sigma}\right)^2$$
$$= \text{constant} - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\epsilon_i(t)^2$$

To bound the influence of outlying observations, we replace $\frac{1}{2}\epsilon_i(t)^2$ by Huber's ρ function:

$$\rho(e) = \frac{1}{2}e^2, \qquad \text{if } |e| \le c$$
$$= c|e| - \frac{1}{2}c^2, \quad \text{if } |e| > c$$

where c is some constant. We use $c = 1.345$; it produces 95% efficiency when the errors are normal and still offers protection against outliers. The robust log-likelihood becomes

$$\text{constant} - \frac{1}{2}\kappa\log\sigma^2 - \frac{1}{2}\kappa\rho(\epsilon_i(t))$$

with the consistent correction factor $\kappa = E[e\frac{\partial}{\partial e}\rho(e)]$ (Huggins, 1993).

For both methods, the maximization of pseudo-likelihood with respect to the unknown parameters can be carried out by the Newton Raphson method. We calculate the conditional expectation in equations (5) and (6) by numerical integration over the unobserved outcomes using the 20-point Gauss Hermite quadrature (Abramowitz et al., 1972). To choose the initial values for maximization by the Newton Raphson method, we first conduct a multiple linear regression to regress Y on covariate X to obtain the initial values for β and σ. Then a logistic regression is used to obtain the initial value for α_w, which estimates the association between missing indicators (R) and W. The initial value for the α_y is 0. These estimates are biased, but they work well as initial values for maximization. This procedure is implemented in a short R program. We intend to make it available as an R package and the reader can write to the first author for the code.

3. SIMULATION STUDY

In this section, we carry out simulation studies to evaluate the robust property and efficiency of the proposed methods. The set-up of the simulation is based on the selection model. The longitudinal data are generated by the linear mixed effects model:

$$(7) \qquad Y_i(t) = \beta_0 + \beta_1 X_i + \beta_2 t + \gamma_i + e_i(t)$$

where $Y_i(t)$ is the outcome value of the ith subject at time t, X_i is the treatment indicator, t is the time, $i = 1, \ldots, n = 200$, $t = 1, \ldots, K = 5$. The sample size is $n = 200$, and each subject has at most $K = 5$ visits. X_i is generated according to a Bernoulli distribution with probability 0.5. β's are the regression coefficients; β_0 is the intercept, β_1 is the treatment effect, and β_2 is the slope of the time trend. γ_i is the ith subject's random effect and $e_i(t)$ is the measurement error for the ith subject's outcome at time t. The missing data are generated by the logistic model,

$$(8) \quad \text{Prob}(R_i(t) = 1) = \frac{\exp(\alpha_0 + \alpha_1 Y_i(t) + \alpha_2 t)}{1 + \exp(\alpha_0 + \alpha_1 Y_i(t) + \alpha_2 t)}$$

where $R_i(t)$ is the response indicator; $R_i(t) = 1$ if $Y_i(t)$ is observed, and 0 if $Y_i(t)$ is missing. The probability that $Y_i(t)$ is missing is determined by the (unobserved) outcome value $Y_i(t)$ and time t. If $\alpha_1 \neq 0$, the missing probability depends on $Y_i(t)$ and the above model generates MNAR data.

For each simulated dataset, we consider 4 methods to estimate the parameters of interest β: maximum pseudo-likelihood estimation based on t distribution model (t model), maximum pseudo-likelihood estimation based on robust normal likelihood (robust normal model), maximum pseudo-likelihood estimation based on normal distribution model (normal model) (Troxel et al., 1998) and linear mixed effects model (LME model) (Laird et al., 1982). The first three models allow MNAR, while the LME model uses model (7) and ignores the missing mechanism in (8).

3.1 Influence of a single outlier

We first illustrate the robust property of the proposed methods by examining the influence of one outlying observation, using a similar approach by Pinheiro et al. (2001). One data set is generated based on normal random effects and normal measurement errors based on (7) and (8) with

$$(9) \qquad \gamma_i \sim N(0, \sigma_\gamma^2) \quad \text{and} \quad e_i(t) \sim N(0, \sigma_e^2).$$

The simulation parameters are $(\beta_0, \beta_1, \beta_2, \sigma_\gamma, \sigma_e, \alpha_0, \alpha_1, \alpha_2) = (1, 0.2, 1, \sqrt{0.5}, \sqrt{0.5}, -4, -0.5, -0.1)$, and this set-up generates about 10% missing data. The sample size (n) is 200, and the number of visits (K) is 5. We arbitrarily replace one observed data point $Y_i(t)$ by $Y_i(t) + \Delta$ to generate an outlying observation. Δ varies between -5 and 5, while the standard deviation of $Y_i(t)$ conditional on X_i and t is
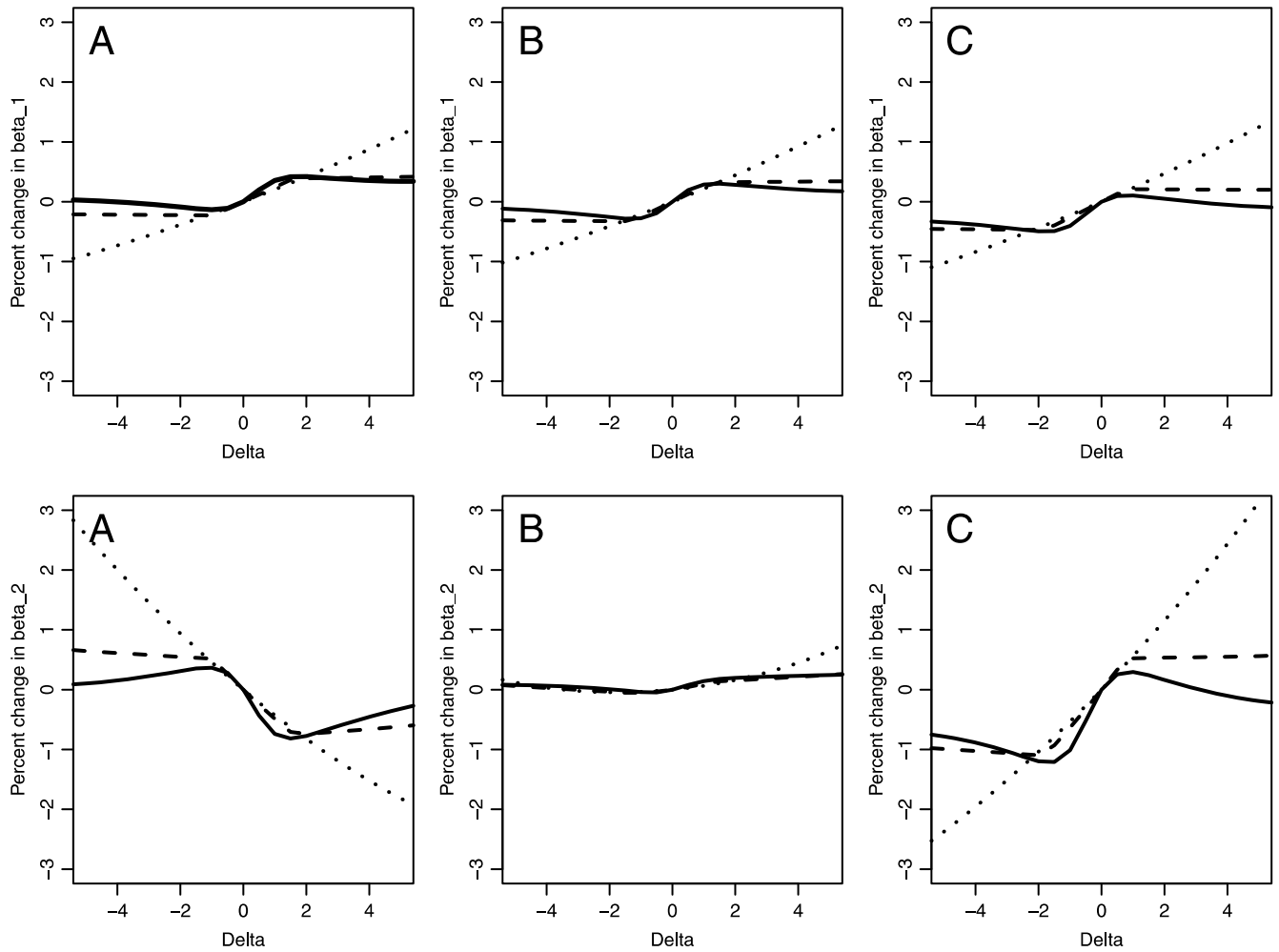
Figure 1. The influence of single outlier on the estimates of treatment effect β_1 and slope β_2. The upper 3 plots of A, B and C give the relative change of $\hat{\beta}_1$ when adding Δ to an observation at an early time ($t = 1$), middle time ($t = 3$), and later time ($t = 5$). The lower 3 plots give the corresponding relative change of $\hat{\beta}_2$. The solid line is the relative change based on the t model, the dashed line is the relative change based on the robust normal model, and the dotted line is the relative change based on the normal model.

$\sqrt{\sigma_\gamma^2 + \sigma_e^2} = 1$. We are interested in the relative change of β estimates. The relative change of β estimates is calculated by $(\hat{\beta}(\Delta) - \hat{\beta})/\hat{\beta}$, where $\hat{\beta}$ is the estimate based on the original data and $\hat{\beta}(\Delta)$ is the estimate based on the new data with an outlying observation.

Figure 1 displays the relative change of treatment effect estimate ($\hat{\beta}_1$) and slope estimate ($\hat{\beta}_2$) based on the t model (solid line) robust normal model (dashed line) and the normal model (dotted line). The upper 3 plots of A, B and C show the relative change of $\hat{\beta}_1$ when adding Δ to an observation at an early time ($t = 1$), middle time ($t = 3$), and later time($t = 5$), respectively. The lower 3 plots give the corresponding relative change of $\hat{\beta}_2$ by Δ. This figure shows that the estimates from the t model and the robust normal model are protected against the influence of an outlying observa-

tion; the estimates based on the t model and robust normal model stay constant as the magnitude of Δ increases. On the other hand, the percent change on the estimates based on the normal model is almost linearly associated with the amount of contamination, and the variance of the estimates also increase sharply as the Δ increases (data not shown). We also experimented with Δ as large as 20, and the estimates based on the t model and the robust model are essentially not affected. For β_1, the impact of Δ is very similar across time points $t = 1$, 3, or 5. Because β_2 is the slope parameter, its change depends on the time point where the outlying observation is located.

3.2 Influence of different outlier patterns

In the second simulation, we generate data sets with various combinations of outlying random effects and outlying

measurement errors in the longitudinal data. Conceptually, outlying random effects represent a proportion of subjects who might behave differently than the rest of the cohort. Outlying measurement errors represent a proportion of observations following a different distribution from the rest of the data. Specifically, in equation (7), the random effects and measurement errors have the following mixture distributions,

$$
(10) \quad
\begin{aligned}
\gamma_i &\sim (1 - p_\gamma)N(0, \sigma_\gamma^2) + p_\gamma N(0, f_\gamma^2 \sigma_\gamma^2) \quad \text{and} \\
e_i(t) &\sim (1 - p_e)N(0, \sigma_e^2) + p_e N(0, f_e^2 \sigma_e^2),
\end{aligned}
$$

where p_γ and p_e correspond to the percentages of random effect outliers and measurement error outliers, and $f_\gamma > 1$ and $f_e > 1$ represent the degree that the outliers deviate from the rest of the observations. The larger f_γ and f_e are, the larger variability the outliers have. A special case $p_\gamma = p_e = 0$ gives normal random effects and normal measurement errors without any outlier.

We are also interested in investigating the impact of MNAR assumption on the proposed methods when this assumption does not hold. The missing data are generated based on the following missing mechanism,

(11)

$$
\text{Prob}(R_i(t) = 1 | Z_i(t)) = \frac{\exp(\alpha_0 + \alpha_1 Z_i(t) Y_i(t) + \alpha_2 t)}{1 + \exp(\alpha_0 + \alpha_1 Z_i(t) Y_i(t) + \alpha_2 t)},
$$

where $Z_i(t)$ a Bernoulli random variable with probability $\text{Prob}(Z_i(t) = 1) = p_m$ and $\text{Prob}(Z_i(t) = 0) = 1 - p_m$. When $p_m = 1$, the missing mechanism is MNAR because the probability of missing always depends on the outcome variable $Y_i(t)$, provided $\alpha_1 \neq 0$. On the other hand when $p_m = 0$, the missing mechanism is MAR. When p_m is between 0 and 1, it generates data by a mixture of MNAR and MAR mechanism.

We vary the combination of $p_m, p_\gamma, p_e, f_\gamma$ and f_e in the simulation and estimate the treatment effects with the 4 models: t model, robust normal model, normal model and LME model. The simulation parameters are $(\beta_0, \beta_1, \beta_2, \sigma_\gamma, \sigma_e, \alpha_0, \alpha_1, \alpha_2) = (1, 0.2, 1, \sqrt{0.5}, \sqrt{0.5}, -4, -0.5, -0.1)$. The sample size is $n = 200$, and each subject has at most $K = 5$ visits. This simulation setup generates about 10% missing data. The LME model assumes MAR missing mechanism, and the normal model, the robust normal model and the t model assume the non-ignorable missingness (MNAR) in equation (8). For each setup, 500 simulations are carried out to compare the bias, the mean square error and to calculate the relative efficiency among these models.

Table 1 gives the simulation summary on the estimation of treatment effect (β_1). We first investigate the bias among 3 missing mechanisms MNAR $(p_m = 1)$, mixture of MNAR and MAR $(p_m = 0.5)$, and MAR $(p_m = 0)$. Overall all models give relatively small bias. Under the MNAR simulation, the LME model appears to produce larger bias on

the estimation of β_1 than the other 3 models that assume non-ignorable missing. When the data are generated under the MAR assumption, the t model appears to have a larger bias than the other 3 models. In theory, the t model and the normal model provide consistent estimators under MNAR when the model and the distribution are correctly specified as shown in Troxel et al. (1998).

We use the ratio of mean square error as a measure of the relative efficiency between two models. R_1 and R_2 give the relative efficiency for the normal vs. the t model and the LME model vs. the t model. R_3 and R_4 give the relative efficiency for the normal model vs. the robust normal model and the LME vs. the robust normal model. If the relative efficiency is larger than one, it suggests that the t model or the robust normal model has a smaller mean square error and is more efficient than the LME or the normal model. In Table 1, we observe that the robust inference based on the t model and the robust normal model offer comparable efficiency comparing to the normal or the LME model when the outliers are of moderate variability $(f_e = 2$ or $f_\gamma = 2)$ or there is no outlier $(p_e = p_\gamma = 0)$. When there exist outliers with large variability $(f_e = 4$ or $f_\gamma = 4)$, the t model and the robust normal model provide better efficiency across different missing mechanism of MNAR, mixture of MNAR and MAR, and MAR. The advantage of robust inference is more pronounced when the outliers are random effect outliers than measurement error outliers. The efficiency also increases as the proportion of outliers increases. The performance of the t model and the robust normal model are similar. The t model works slightly better when the outlier proportion is high and variability is large. One the other hand, when there is no outlier, the robust model has better efficiency than the t model and it maintains about 95% efficiency compared to the normal model or the LME model.

Table 2 summarizes the simulation results for the slope of time trend β_2. Overall, the bias is small and similar among 4 models. The LME model has a slightly larger bias with MNAR data. We still observe that the t model and the robust normal model offer better efficiency when the outlier proportion is high and variability is large, particularly with measurement error outliers. However, when there is no outlier or the outliers have moderate variability, the normal model and the LME model appear to have better efficiency, and the robust normal model also outperforms the t model. The t model and the robust model have more pronounced advantage in estimating β_2 with measurement error outliers than random effect outliers.

4. SCLERODERMA LUNG STUDY DATA ANALYSIS

In this section, we demonstrate the use of the proposed methods to handle non-monotonic and non-ignorable missingness in the analysis of the Scleroderma Lung Study

Table 1. Simulation studies to compare bias, mean square error(MSE) on β_1 between the t model, the robust normal model, the normal model and the linear mixed effects (LME) model. The relative efficiency (RE) is calculated for
$$RE_1 = \frac{MSE(normal\ model)}{MSE(t\ model)},\ RE_2 = \frac{MSE(LME)}{MSE(t\ model)},\ RE_3 = \frac{MSE(normal\ model)}{MSE(robust\ model)},\ RE_4 = \frac{MSE(LME)}{MSE(robust\ model)}$$

MNAR, $p_m = 1$				Bias($\times 10^{-3}$)				MSE($\times 10^{-3}$)							
p_γ	p_e	f_γ	f_e	t	robust	normal	LME	t	robust	normal	LME	RE_1	RE_2	RE_3	RE_4
0	0	1	1	0.23	3.02	3.99	9.86	12.67	12.19	11.84	11.77	0.93	0.93	0.97	0.97
0.05	0	2	1	8.68	6.26	5.36	17.34	13.33	12.46	12.79	12.76	0.96	0.96	1.03	1.02
0.1	0	2	1	2.87	5.61	6.90	4.90	15.98	15.16	15.65	15.68	0.98	0.98	1.03	1.03
0.25	0	2	1	4.85	4.37	2.86	7.57	16.73	17.20	18.89	18.59	1.13	1.11	1.10	1.08
0.05	0	4	1	0.54	4.21	6.11	17.77	15.45	14.92	18.65	18.92	1.21	1.22	1.25	1.27
0.1	0	4	1	7.02	6.09	9.54	18.52	16.03	16.12	24.36	25.00	1.52	1.56	1.51	1.55
0.25	0	4	1	25.86	27.89	41.91	25.51	21.9	26.46	42.39	45.41	1.94	2.07	1.6	1.72
0	0.05	1	2	6.79	7.64	7.53	6.86	13.71	13.12	12.93	12.73	0.94	0.93	0.99	0.97
0	0.1	1	2	3.47	1.22	1.38	12.01	13.28	12.97	12.11	11.71	0.91	0.88	0.93	0.9
0	0.25	1	2	0.40	0.81	1.22	18.14	15.16	14.25	14.06	14.16	0.93	0.93	0.99	0.99
0	0.05	1	4	6.03	5.04	4.93	21.80	14.96	14.59	16.70	16.30	1.12	1.09	1.14	1.12
0	0.1	1	4	4.21	1.52	2.30	13.03	13.93	13.36	14.33	13.98	1.03	1.00	1.07	1.05
0	0.25	1	4	26.79	25.08	40.49	46.98	14.77	15.44	20.11	20.47	1.36	1.39	1.3	1.33
$p_m = 0.5$				Bias($\times 10^{-3}$)				MSE($\times 10^{-3}$)							
p_γ	p_e	f_γ	f_e	t	robust	normal	LME	t	robust	normal	LME	RE_1	RE_2	RE_3	RE_4
0	0	1	1	8.76	2.26	3.05	3.27	13.36	11.97	11.35	11.33	0.85	0.85	0.95	0.95
0.05	0	2	1	11.33	4.86	4.45	1.84	13.37	13.02	13.42	13.48	1.00	1.01	1.03	1.03
0.1	0	2	1	13.22	9.35	9.31	2.97	15.12	14.52	15.04	14.90	1.00	0.99	1.04	1.03
0.25	0	2	1	5.93	8.29	8.35	13.51	20.08	19.86	22.19	22.16	1.10	1.10	1.12	1.12
0.05	0	4	1	3.54	5.48	6.71	10.03	14.86	15.03	19.21	19.91	1.29	1.34	1.28	1.32
0.1	0	4	1	16.88	19.00	26.91	27.56	15.58	16.36	23.79	24.35	1.53	1.56	1.45	1.49
0.25	0	4	1	3.46	4.47	13.73	1.00	23.74	27.87	43.50	45.15	1.83	1.9	1.56	1.62
0	0.05	1	2	7.57	1.83	1.04	6.01	13.58	13.03	12.78	12.72	0.94	0.94	0.98	0.98
0	0.1	1	2	8.64	3.60	3.29	3.23	13.95	13.13	12.78	12.53	0.92	0.90	0.97	0.95
0	0.25	1	2	1.40	0.51	0.96	5.47	13.95	12.94	12.94	13.08	0.93	0.94	1.00	1.01
0	0.05	1	4	11.06	8.81	8.15	12.8	14.30	14.27	16.92	16.86	1.18	1.18	1.19	1.18
0	0.1	1	4	6.65	6.90	7.11	1.35	12.86	12.33	13.04	12.93	1.01	1.01	1.06	1.05
0	0.25	1	4	20.83	20.98	29.08	26.51	16.62	16.92	21.16	20.99	1.27	1.26	1.25	1.24
MAR, $p_m = 0$				Bias($\times 10^{-3}$)				MSE($\times 10^{-3}$)							
p_γ	p_e	f_γ	f_e	t	robust	normal	LME	t	robust	normal	LME	RE_1	RE_2	RE_3	RE_4
0	0	1	1	11.73	5.18	6.41	3.00	13.9	12.26	12.10	12.10	0.87	0.87	0.99	0.99
0.05	0	2	1	8.60	2.24	1.80	2.03	14.32	14.07	14.53	14.43	1.01	1.01	1.03	1.03
0.1	0	2	1	13.19	7.43	7.01	2.61	15.31	14.78	15.33	15.36	1.00	1.00	1.04	1.04
0.25	0	2	1	22.42	14.02	15.64	10.21	18.30	17.52	19.01	18.80	1.04	1.03	1.09	1.07
0.05	0	4	1	16.71	11.92	13.06	8.30	15.65	15.17	20.46	20.07	1.31	1.28	1.35	1.32
0.1	0	4	1	14.34	9.18	8.72	2.99	17.50	18.72	30.66	30.47	1.75	1.74	1.64	1.63
0.25	0	4	1	3.63	3.34	5.08	1.55	25.66	30.69	51.32	50.89	2.00	1.98	1.67	1.66
0	0.05	1	2	8.42	2.76	1.90	2.07	13.07	12.59	12.22	12.28	0.93	0.94	0.97	0.98
0	0.1	1	2	9.00	1.29	1.38	3.08	13.26	12.81	12.71	12.65	0.96	0.95	0.99	0.99
0	0.25	1	2	8.76	1.71	0.57	3.81	12.88	11.83	12.20	12.15	0.95	0.94	1.03	1.03
0	0.05	1	4	2.08	0.86	2.68	1.09	13.79	13.29	15.32	15.32	1.11	1.11	1.15	1.15
0	0.1	1	4	10.42	4.41	5.04	1.31	13.39	12.78	13.78	13.76	1.03	1.03	1.08	1.08
0	0.25	1	4	8.24	9.69	11.06	9.91	15.41	15.96	19.59	19.67	1.27	1.28	1.23	1.23

(Tashkin et al., 2006). The Scleroderma Lung Study is a multi-center placebo-control double-blind randomized study to evaluate the effects of oral cyclophosphamide (CYC) on lung function and other health-related symptoms in patients with evidence of active alveolitis and scleroderma-related interstitial lung disease. In this study, eligible participants received either daily oral cyclophosphamide or matching placebo for 12 months, followed by another year of follow-up without study medication. The primary end point of the study is the forced vital capacity (FVC, expressed as a percentage of the predicted value), which is measured at baseline and at three-month intervals throughout the study. One hundred and fifty eight eligible patients underwent randomization, and about 15% of them dropped out of the study
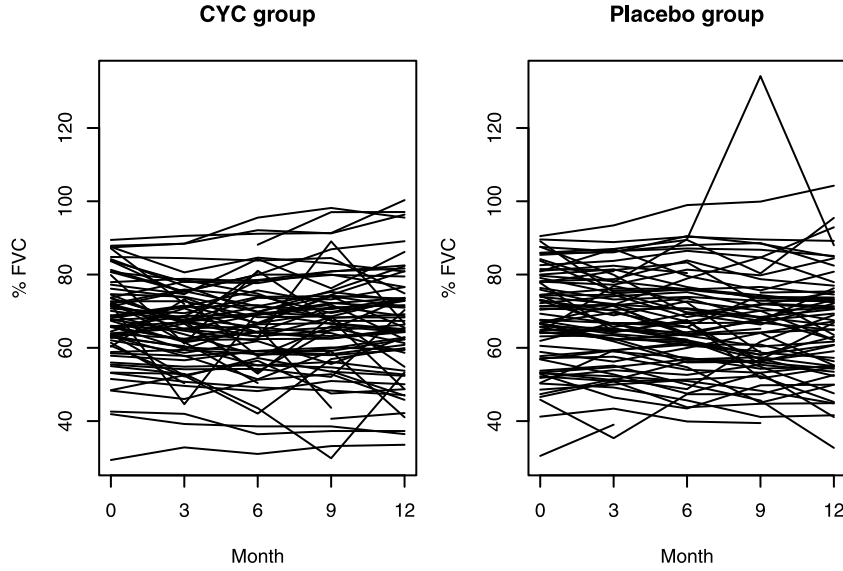
Figure 2. FVC overtime in the placebo and CYC groups. In the Scleroderma lung study, the forced vital capacity (FVC, expressed as a percentage of the predicted value) is measured at baseline and at three-month intervals throughout the study.

before 12 months. About 30% of dropouts are due to death and treatment failures. Intermittent missed visits also occurred during the course of the study. It is likely that the missing data are due to the ineffectiveness of treatment and related to the outcome of interest.

Here we present the analysis of the primary endpoint of FVC. Figure 2 gives the plot of FVC over time for each study group. Most subjects have small variation in FVC over time. However, a few observations show considerable changes between visits. It suggests the possibility of outlying observations. The outcome model includes covariates of time, baseline FVC (FVC_0), baseline maximum fibrosis ($MAXFIB_0$), cyclophosphamide (CYC), and the interactions between treatment and baseline FVC, baseline maximal fibrosis and time.

$$FVC = \beta_0 + \beta_1 Time + \beta_2 FVC_0 + \beta_3 MAXFIB_0$$
$$+ \beta_4 CYC + \beta_5 CYC \times FVC_0$$
$$+ \beta_6 CYC \times MAXFIB_0 + \beta_7 CYC \times Time + \epsilon$$

We also specify the missing data model that the missing probability is a function of FVC and time, using a logistic model.

$$Pr(R = 0) = [1 + \exp(\alpha_0 + \alpha_1 FVC + \alpha_2 Time)]^{-1}$$

Table 3 gives the analyses of the outcome model based on the t model, robust normal model, normal model and LME model. Comparing the t model, robust normal model and normal model, it appears that normal model is more sensitive to outlying observations and generates larger estimated coefficients and larger standard deviation than the other two models. For example, the estimated β_1 is -0.33, -0.65, and

-1.15 for t model, robust normal model and normal model. Similar trends are also observed for β_3, β_4, and β_6.

Because this is a randomized trial, the assignment of CYC is independent of baseline characteristics. We can derive the overall treatment effect at time m:

$$E(FVC|CYC = 1, Time = m)$$
$$- E(FVC|CYC = 0, Time = m)$$
$$= \int \big[E(FVC|CYC = 1, Time = m, FVC_0, MAXFIB_0)$$
$$- E(FVC|CYC = 0, Time = m, FVC_0, MAXFIB_0) \big]$$
$$\times df(FVC_0, MAXFIB_0)$$
$$\approx 1/n \sum_{i=1}^{n} Eff_i$$

where Eff_i is the treatment(CYC) effect for the ith subject at time m

$$Eff_i$$
$$= E(FVC_i|CYC = 1, Time = m, FVC_{0i}, MAXFIB_{0i})$$
$$- E(FVC_i|CYC = 0, Time = m, FVC_{0i}, MAXFIB_{0i}).$$

Based on the t model the overall treatment effect at 12 months is estimated by 2.41(%) with the standard deviation of 1.18, and the p-value $= 0.041$. The robust normal model estimates similar treatment effect 2.50 with standard deviation of 2.33. The normal model and LME model give smaller treatment effect estimates of 1.80 and 1.47. We also observe that the normal model gives the largest variance estimate. It suggests that the t-model may be more sensitive in detecting the treatment effect. Alternatively, we can use the Wald's test to examine the hypothesis of $\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

Table 2. Simulation studies to compare bias, mean square error(MSE) on β_2 between the t model, the robust normal model, the normal model and the linear mixed effects (LME) model. The relative efficiency (RE) is calculated for
$$RE_1 = \frac{MSE(normal\ model)}{MSE(t\ model)},\ RE_2 = \frac{MSE(LME)}{MSE(t\ model)},\ RE_3 = \frac{MSE(normal\ model)}{MSE(robust\ model)},\ RE_4 = \frac{MSE(LME)}{MSE(robust\ model)}$$

MNAR, $p_m = 1$				Bias($\times 10^{-3}$)				MSE($\times 10^{-3}$)							
p_γ	p_e	f_γ	f_e	t	robust	normal	LME	t	robust	normal	LME	RE_1	RE_2	RE_3	RE_4
0	0	1	1	0.57	0.78	0.53	4.17	0.41	0.32	0.29	0.28	0.71	0.68	0.92	0.88
0.05	0	2	1	1.28	1.15	1.92	4.71	0.42	0.34	0.31	0.30	0.75	0.72	0.93	0.89
0.1	0	2	1	1.84	0.33	0.53	3.66	0.42	0.32	0.29	0.28	0.68	0.66	0.91	0.88
0.25	0	2	1	3.52	1.51	1.96	3.92	0.50	0.40	0.37	0.31	0.73	0.61	0.92	0.77
0.05	0	4	1	2.60	2.72	2.00	4.62	0.40	0.33	0.32	0.31	0.80	0.76	0.97	0.92
0.1	0	4	1	5.07	3.64	3.08	4.63	0.46	0.40	0.40	0.33	0.86	0.71	1.01	0.83
0.25	0	4	1	8.90	8.17	13.78	3.54	0.53	0.45	0.65	0.29	1.23	0.55	1.44	0.64
0	0.05	1	2	0.51	0.15	0.48	4.56	0.37	0.31	0.30	0.30	0.82	0.81	0.97	0.95
0	0.1	1	2	2.00	0.92	0.70	4.51	0.45	0.38	0.39	0.39	0.86	0.85	1.02	1.01
0	0.25	1	2	2.92	1.62	2.22	6.83	0.51	0.45	0.46	0.50	0.91	0.99	1.02	1.11
0	0.05	1	4	4.03	3.58	3.83	8.92	0.46	0.43	0.63	0.66	1.36	1.42	1.47	1.53
0	0.1	1	4	3.73	2.34	1.12	6.00	0.38	0.36	0.46	0.48	1.21	1.27	1.29	1.36
0	0.25	1	4	9.45	8.23	12.56	13.33	0.69	0.73	1.27	1.27	1.84	1.84	1.73	1.73
$p_m = 0.5$				Bias($\times 10^{-3}$)				MSE($\times 10^{-3}$)							
p_γ	p_e	f_γ	f_e	t	robust	normal	LME	t	robust	normal	LME	RE_1	RE_2	RE_3	RE_4
0	0	1	1	2.36	0.89	0.74	2.08	0.42	0.31	0.28	0.27	0.67	0.64	0.91	0.87
0.05	0	2	1	1.79	1.16	0.98	2.50	0.46	0.34	0.31	0.28	0.67	0.62	0.89	0.82
0.1	0	2	1	1.72	0.85	1.03	2.04	0.41	0.32	0.28	0.26	0.68	0.63	0.88	0.82
0.25	0	2	1	2.53	1.03	1.59	2.24	0.48	0.35	0.30	0.28	0.63	0.59	0.85	0.8
0.05	0	4	1	3.20	2.09	2.03	2.56	0.37	0.31	0.33	0.29	0.88	0.77	1.05	0.92
0.1	0	4	1	3.44	2.46	2.96	2.29	0.40	0.32	0.32	0.29	0.80	0.73	0.98	0.89
0.25	0	4	1	4.44	4.83	9.01	1.28	0.48	0.42	0.51	0.32	1.07	0.67	1.21	0.76
0	0.05	1	2	1.33	0.43	0.55	2.30	0.47	0.36	0.34	0.33	0.71	0.69	0.95	0.93
0	0.1	1	2	1.75	0.44	0.29	1.83	0.47	0.36	0.35	0.34	0.75	0.71	0.98	0.94
0	0.25	1	2	4.45	2.07	2.33	3.80	0.57	0.46	0.48	0.47	0.84	0.83	1.05	1.04
0	0.05	1	4	2.88	2.12	3.08	3.84	0.46	0.44	0.63	0.63	1.38	1.37	1.44	1.43
0	0.1	1	4	2.82	1.58	0.97	2.23	0.38	0.34	0.47	0.46	1.22	1.20	1.36	1.35
0	0.25	1	4	5.94	5.75	8.31	6.36	0.67	0.76	1.32	1.29	1.96	1.92	1.73	1.69
MAR, $p_m = 0$				Bias($\times 10^{-3}$)				MSE($\times 10^{-3}$)							
p_γ	p_e	f_γ	f_e	t	robust	normal	LME	t	robust	normal	LME	RE_1	RE_2	RE_3	RE_4
0	0	1	1	0.23	0.14	0.32	0.09	0.37	0.28	0.26	0.25	0.70	0.68	0.94	0.91
0.05	0	2	1	2.07	0.66	0.25	0.42	0.34	0.28	0.26	0.25	0.76	0.75	0.92	0.90
0.1	0	2	1	0.60	1.58	1.21	1.42	0.34	0.28	0.26	0.25	0.77	0.74	0.94	0.90
0.25	0	2	1	1.64	0.75	1.19	0.97	0.40	0.32	0.31	0.28	0.77	0.71	0.95	0.87
0.05	0	4	1	0.60	0.21	0.30	0.52	0.38	0.30	0.28	0.27	0.74	0.70	0.93	0.87
0.1	0	4	1	2.34	1.22	1.50	1.24	0.35	0.28	0.27	0.25	0.77	0.72	0.95	0.88
0.25	0	4	1	1.27	1.81	0.59	1.45	0.38	0.30	0.29	0.24	0.75	0.63	0.95	0.78
0	0.05	1	2	0.28	0.38	0.03	0.05	0.37	0.31	0.30	0.30	0.80	0.80	0.95	0.95
0	0.1	1	2	2.47	1.49	1.52	1.29	0.39	0.31	0.32	0.31	0.82	0.80	1.04	1.01
0	0.25	1	2	1.07	0.12	0.00	0.38	0.46	0.41	0.43	0.42	0.95	0.93	1.06	1.04
0	0.05	1	4	0.60	0.58	0.34	0.78	0.46	0.46	0.71	0.71	1.55	1.54	1.54	1.53
0	0.1	1	4	2.03	0.79	0.71	0.48	0.39	0.35	0.46	0.45	1.17	1.14	1.32	1.29
0	0.25	1	4	0.77	0.64	2.35	1.04	0.58	0.66	1.17	1.16	2.03	2.02	1.78	1.77

for overall treatment effect. All 4 methods show clear evidence of treatment difference. These results are in agreement with the analysis in Tashkin et al. (2006). The t model, robust normal model and normal model also give significant p-values for FVC(α_1) and Time (α_2) on their associations with missing data. It suggests that, if the models are specified correctly, the missing mechanism is likely MNAR.

5. CONCLUSIONS

We present two methods to provide robust inference for longitudinal data analysis with non-ignorable and non-monotonic missing values. These two methods are robust in two ways: first, the t-distribution model and the robust normal likelihood reduce the influence of outlying observa-

Table 3. *The comparison of t model, robust normal model, normal model and linear mixed effects model for FVC in the Scleroderma study, using first 12 months data*

β	t model Est(Std)	p-value	robust model Est(Std)	p-value	normal model Est(Std)	p-value	LME model Est(Std)	p-value
β_1(Time)	$-0.33(0.07)$	<0.001	$-0.65(0.16)$	<0.001	$-1.15(0.24)$	<0.001	$-0.27(0.08)$	<0.001
β_2(FVC$_0$)	$0.95(0.03)$	<0.001	$0.93(0.07)$	<0.001	$0.82(0.12)$	<0.001	$0.92(0.04)$	<0.001
β_3(MAXFIB$_0$)	$-1.28(0.29)$	<0.001	$-2.14(0.57)$	<0.001	$-3.22(1.15)$	0.005	$-1.32(0.43)$	0.003
β_4(CYC)	$-0.74(0.25)$	0.003	$-1.50(0.58)$	0.010	$-2.00(1.43)$	0.161	$-0.77(0.89)$	0.387
β_5(CYC×FVC$_0$)	$0.05(0.05)$	0.303	$0.04(0.09)$	0.661	$0.04(0.19)$	0.840	$0.07(0.05)$	0.208
β_6(CYC×MAXFIB$_0$)	$1.77(0.47)$	<0.001	$3.16(1.10)$	0.004	$5.47(2.18)$	0.012	$1.48(0.62)$	0.018
β_7(CYC×Time)	$0.26(0.10)$	0.008	$0.33(0.19)$	0.080	$0.32(0.33)$	0.341	$0.19(0.11)$	0.102
overall effect at 12M	$2.41(1.18)$	0.041	$2.50(2.33)$	0.28	$1.80(4.43)$	0.684	$1.47(0.95)$	0.123
$\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$		<0.001		<0.001		0.011		0.055

tions on the parameter estimation and, second, the pseudo-likelihood approach does not impose any constraint on the joint distribution of data. We demonstrate that an analysis with linear mixed effects models can potentially produce larger bias when the missing mechanism is MNAR. Both the t model and the robust normal model have better efficiency than the normal model or linear mixed effects model, when there is large percentage of long tailed outlying observations. However, when there are no or some outlying observations with moderate variability, robust normal and t model can be less efficient than the normal model or linear mixed effects model.

One marginal outcome model and one missing mechanism model are specified to analyse the Scleroderma data and in simulations study. In reality, subjects may miss visits or be lost to follow-up for different reasons. For example, possible reasons for study dropout in the Scleroderma study include death, treatment failure, loss to follow-up and etc. A potential improvement of our model is to incorporate different missing mechanism models for different causes of missing data to better describe the missing mechanism. For example, Elashoff et al. (2007), Elashoff et al. (2008) and Li et al. (2009) specify different survival functions for time to different dropout causes. A clear and detailed documentation of missing reasons and dropout causes is imperative for such analysis. We use a logistic regression model to estimate the missingness probability for non-monotonic missing data. However, the missing data pattern in a clinical study can include both monotonic and non-monotonic missingness. In particular, intermittent missingness can occur for all alive subjects, while missing data due to death are always monotonic. A missing mechanism model that distinguishes non-monotonic and monotonic might be more appropriate.

Our t model has a fixed 3 degrees of freedom. It is possible to consider the degrees of freedom as a parameter of the model and estimate it based on the data. However, we observe that our t model with fixed 3 degrees of freedom provides more than 85% efficiency in estimating the treatment effect even when the data are normally distributed. Estimating the degrees of freedom generally involves intensive

computation (Pinheiro et al., 2001) and the improvement in efficiency might be limited.

Finally, the selection model provides a natural way to model the longitudinal process and missing mechanism simultaneously and has the advantage of easy interpretation. It, however, has the drawback of non-identifiability that the distribution of missing data and the missing mechanism can not be verified based on the observed data. Fortunately, we can often bring in knowledge and assumptions that are external to the data to provide evidence for missing mechanism (Verbeke et al., 2008). Therefore the non-ignorable missing data analysis must be carried out with great care. It also implies the importance of collecting information on the causes of missing data in a clinical study to determine possible missing mechanism. In our Scleroderma example, many missing data are due to lack of treatment efficacy (for example, death or treatment failure), and MNAR seems to be a reasonable assumption. In general, local sensitivity analysis can be performed to evaluate the effects of non-ignorability (Verbeke et al., 2001; Ma et al., 2005). It is also desirable to investigate the sensitivity of different parametrization of missing data models and its implication on the robust property of estimates (Rizopoulos et al., 2008).

APPENDIX

Here we demonstrate the consistency for the pseudo-likelihood estimator of β. We denote $f_t(Y)$ and $f_t(R|Y)$ as the density function based on t or robust normal model and its associated missing mechanism model, and $f(Y)$ and $f(R|Y)$ as the density function of the true marginal model of response Y and the true missing mechanism. The parameter of interest is β in the longitudinal model and we use $f'_t(Y)$ to represent $\frac{\partial}{\partial \beta} f_t(Y)$.

To use the method of moment theory (Hansen, 1982) to demonstrate the consistency of t model estimators, we need to first evaluate the expected value of pseudo-likelihood score equation (equation (3)): $E(S(\beta)) = \sum_{i=1}^{n} S_i(\beta) =$

$\sum_{i=1}^{n} \sum_{t=1}^{K} S_i(t)(\beta)$, and

$$(12) \quad E(S_i(t)(\beta)) = P(R(t)_1 = 0)E(S(t)_i|R(t)_i = 0) \\ + P(R(t)_i = 1)E(S(t)_i|R(t)_i = 1).$$

We first examine $E(S(t)_i|R(t)_i = 0)$. For notational simplicity, we will drop the time indicator t and use S_i, R_i and Y_i to represent $S_i(t)$, $R_i(t)$ and $Y_i(t)$ such that

$$E(S_i|R_i = 0) \\ = \int_{Y_i} \left[\frac{\partial}{\partial \beta} \log f_t(Y_i) + \log f_t(R_i = 0|Y_i) \right] f(Y_i|R_i = 0)dY_i \\ = \int_{Y_i} \left[\frac{\partial}{\partial \beta} log f_t(Y_i) \right] f(Y_i|R_i = 0)dY_i \\ = \int_{Y_i} \frac{f_t'(Y_i)}{f_t(Y_i)} \frac{f(Y_i)f(R_i = 0|Y_i)}{P(R_i = 0)} dY_i.$$

The first term of equation (12) becomes

$$(13)$$
$$P(R_1 = 0)E(S_i|R_i = 0) = \int_{Y_i} f_t'(Y_i) \frac{f(Y_i)}{f_t(Y_i)} f(R_i = 0|Y_i)dY_i.$$

We can express $E(S_i|R_i = 1)$ as

$$E(S_i|R_i = 1) \\ = \int_{Y_i} \left[\frac{\partial}{\partial \beta} log \int_{Y_i} f_t(Y_i)f_t(R_i = 1|Y_i)dY_i \right] f(Y_i|R_i = 1)dY_i \\ = \int_{Y_i} \left[\frac{\int_{Y_i} f_t'(Y_i)f_t(R_i = 1|Y_i)dY_i}{\int_{Y_i} f_t(Y_i)f_t(R_i = 1|Y_i)dY_i} \right] f(Y_i|R_i = 1)dY_i \\ = \frac{\int_{Y_i} f_t'(Y_i)f_t(R_i = 1|Y_i)dY_i}{\int_{Y_i} f_t(Y_i)f_t(R_i = 1|Y_i)dY_i} \\ = \frac{\int_{Y_i} f_t'(Y_i)f_t(R_i = 1|Y_i)dY_i}{P_t(R_i = 1)}$$

where $P_t(R_i = 1) = \int_{Y_i} f_t(Y_i)f_t(R_i = 1|Y_i)dY_i$ represents the marginal probability of missing data based on t or robust normal model.

If $P_t(R_i = 1) \approx P(R_i = 1)$,

$$(14)$$
$$P(R_1 = 1)E(S_i|R_i = 1) \approx \int_{Y_i} f_t'(Y_i)f_t(R_i = 1|Y_i)dY_i.$$

Combining (13) and (14),

$$E(S_i(\beta)) \\ = \int_{Y_i} f_t'(Y_i) \frac{f(Y_i)}{f_t(Y_i)} f(R_i = 0|Y_i)dY_i \\ + \int_{Y_i} f_t'(Y_i)f_t(R_i = 1|Y_i)dY_i \\ = \int_{Y_i} f_t'(Y_i) \frac{f(Y_i)f(R_i = 0|Y_i) + f_t(Y_i)f_t(R_i = 1|Y_i)}{f_t(Y_i)} dY_i.$$

If the model assumptions hold, that is, $f_t(Y_i) = f(Y_i)$ and $f_t(R_i|Y_i) = f(R_i|Y_i)$, then

$$E(S_i(\beta)) = \int_{Y_i} f_t'(Y_i)dY_i = \frac{\partial}{\partial \beta} \int f_t(Y_i)dY_i = 0,$$

and the proposed model provides consistent estimates for the longitudinal model based on the method of moments theory (Hansen, 1982). We observe that $f_t(Y_i)$ is an even function and $f_t'(Y_i)$ is an odd function. If $f(Y_i)f(R_i = 0|Y_i) + f_t(Y_i)f_t(R_i = 1|Y_i)$ is also an even function then $E(S_i(\beta)) = 0$. Our simulation results suggest that the estimates based on the t-model have negligible bias with various outliers patterns, percentages and missing mechanisms.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions*, 10th ed.

BROWN, E. R. and IBRAHIM, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* **59**, 221–228. MR1987388

CANTONI, E. and RONCHETTI, E. (2001). Robust Inference for Generalized Linear Models. *Journal of the American Statistical Association* **96**, 1022–1030. MR1947250

ELASHOFF, R. M., LI, G., and LI, N. (2007). An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine* **26**, 2813–2835. MR2370939

ELASHOFF, R. M., LI, G., and LI, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64**, 762–771. MR2526626

GHIDEY, W., LESAFFRE, E., and EILERS, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics* **60**, 945–953. MR2133547

GILL, P. (2000). A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine* **19**, 975–987.

GONG, G. and SAMANIEGO, F. (1981). Pseudo maximum-likelihood estimation-theory and applications. *Annals of Statistics* **9**, 861–869. MR0619289

HANSEN, L. (1982). Sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054. MR0666123

HOGAN, J. W., ROY, J. and KORKONTZELOU, C. (2004). Biostatistics tutorial: Handling dropout in longitudinal data. *Statistics in Medicine* **23**, 1455–1497.

HUGGINS, R. (1993). A robust approach to the analysis of repeated measures. *Biometrics* **49**, 715–720. MR1243487

IBRAHIM, J. and MOLENBERGHS, G. (2009). Missing data methods in longitudinal studies: A review. *TEST* **18**, 1–43. MR2495958

Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (2001). Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88**, 551–564. MR1844851

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the t-distribution. *Journal of the American Statistical Association* **84**, 881–896. MR1134486

Li, N., Elashoff, R. M., and Li, G. (2009). Robust joint modeling of longitudinal measurements and competing risks failure time data. *Biometrical Journal* **51**, 19–30. MR2667508

Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear model. *Biometrika* **73**, 13–22. MR0836430

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. MR1925014

Ma, G., Troxel, A., and Heitjan, D. (2005). An index of local sensitivity to nonignorable drop-out in longitudinal modeling. *Statistics in Medicine* **24**, 2129–2150. MR2146923

Parzen, M., Lipsitz, S., Fitzmaurice, G., Ibrahim, J., and Troxel, A. (2006). Pseudo-likelihood methods for longitudinal binary data with non-ignorable missing responses and covariates. *Statistics in Medicine* **25**, 2784–2796. MR2242203

Pinheiro, J., Liu, C., and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t-distribution. *Journal of Computational and Graphical Statistics* **10**, 249–276. MR1939700

Richardson, A. and Welsh, A. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics* **51**, 1429–1439.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika* **95**, 63–71. MR2409715

Schluchter, M. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* **11**, 1861–1870.

Tashkin, D. P., Elashoff, R., Clements, P. J., Goldin, J., Roth, M. D., Furst, D. E., Arriola, E., Silver, R., Strange, C., Bolster, M., Seibold, J. R., Riley, D. J., Hsu, V. M., Varga, J., Schraufnagel, D. E., Theodore, A., Simms, R., Wise, R., Wigley, F., White, B., Steen, V., Read, C., Mayes, M., Parsley, E., Mubarak, K., Connolly, M. K., Golden, J., Olman, M., Fessler, B., Rotheld, N., Metersky, M., and the Scleroderma Lung Study Research Group (2006). Cyclophosphamide versus placebo in scleroderma lung disease. *N. Engl. J. Med.* **354**, 2655–2666.

Troxel, A. B., Lipsitz, S., and Harrington, D. (1998). Marginal models for the analysis of longitudinal measurements with non-ignorable non-monotone missing data. *Biometrika* **85**, 661–672. MR1665826

Tukey, J. (1949). One degree of freedom for non-additivity. *Biometrics* **5**, 232–242.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* **91**, 217–221.

Verbeke, G., Molenberghs, G., Thijs, H., Lesare, E., and Kenward, M. (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics* **57**, 7–14. MR1833286

Verbeke, G., Molenberghs, G., and Beunckens, C. (2008). Formal and informal model selection with incomplete data. *Statistical Science* **23**, 201–218. MR2516820

Wang, C. Y. and Pepe, M. S. (2000). Expected estimating equations to accommodate covariate measurement error. *Journal of the Royal Statistical Society, Series B* **62**, 509-524. MR1772412

Wang, C. Y., Huang, Y., Chao, E. C. and Jeffcoat, M. K. (2008). Expected estimating equations for missing data, measurement error, and misclassification, with application to longitudinal nonignorable missing data. *Biometrics* **64**, 85–95. MR2422822

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25. MR0640163

Wu, M. and Carroll, R. (1988). Estimation and comparison of change in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188. MR0931633

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **57**, 795–802. MR1859815

Chi-hong Tseng
Department of Medicine
School of Medicine
University of California at Los Angeles
USA
E-mail address: ctseng@mednet.ucla.edu

Robert Elashoff
Department of Biomathematics and Biostatistics
School of Medicine and Public Health
University of California at Los Angeles
USA
E-mail address: relashoff@biomath.ucla.edu

Ning Li
Biostatistics Core
Samuel Oschin Comprehensive Cancer Institute
Cedars-Sinai Medical Center
USA
E-mail address: Ning.Li@cshs.org

Gang Li
Department of Biostatistics
School of Public Health
University of California at Los Angeles
USA
E-mail address: vli@ucla.edu