JIRSS (2012)
Vol. 11, No. 2, pp 101-129

A Non-Random Dropout Model for Analyzing Longitudinal Skew-Normal Response

T. Baghfalaki[1], M. Ganjali[1], M. Khounsiavash[2]

[1]Department of Statistics, Shahid Beheshti University, Tehran, Iran.
[2]Islamic Azad University, Qazvin Branch, Qazvin, Iran.

Abstract. In this paper, multivariate skew-normal distribution is employed for analyzing an outcome based dropout model for repeated measurements with non-random dropout in skew regression data sets. A probit regression is considered as the conditional probability of an observation to be missing given outcomes. A simulation study of using the proposed methodology and comparing it with a semi-parametric method, GEE, is provided. The standardized bias is used for comparison of different approaches. Furthermore, for investigation of efficiency of the methodology two applications are analyzed where observed information matrix is used to find the variances of the parameter estimates. In one of the applications a sensitivity analysis is also performed to investigate the change on the response model's parameter estimates due to perturbation of drop-out model's parameter of interest.

Keywords. Dropout, generalized estimating equations (GEE), longitudinal data, observed information matrix, selection model, Skew-Normal distribution.

MSC: Primary: 62H12; Secondary: 62P10.

T. Baghfalaki (tbaghfalaki@yahoo.com), M. Ganjali(✉)(m-ganjali@sbu.ac.ir), M. Khounsiavash (mfsiavash@gmail.com)
Received: February 7, 2011; Accepted: August 13, 2012

1 Introduction

The skew-normal is a class of distributions that includes the normal one as a special case. In particular, a random variable Z is said to be skew-normal with shape parameter λ, written $Z \sim SN(\lambda)$, if its density function at z is

$$f(z;\lambda) = 2\phi(z)\Phi(\lambda z) \qquad z, \lambda \in R, \tag{1}$$

where ϕ and Φ denote the $N(0,1)$ density and distribution functions, respectively. We note that, when $\lambda = 0$, Z is a standard normal variable, while otherwise the sign of λ gives the sign of the skewness.

In practice, it is common to work with a location and scale transformation $Y = \mu + \sigma Z$ with $\mu \in R$ and $\sigma > 0$. Hence, the density for the random variable Y, written $Y \sim SN(\mu, \sigma, \lambda)$, is

$$f(y;\ \lambda,\ \mu,\ \sigma) = \frac{2}{\sigma}\phi(\frac{y-\mu}{\sigma})\Phi(\lambda\frac{y-\mu}{\sigma}). \tag{2}$$

Liseo and Loperfido (2004) showed that, in standard form of skew-normal distribution when all the observations have the same sign, MLE is infinite and the probability of this to occur is given by $[(0.5-\arctan\lambda)^n + (0.5+\arctan\lambda)^n]$. There are some methods for solving this problem which are proposed by Sartori (2006) and Bayes and Branco (2007).

A systematic treatment of the skew-normal distribution has been given in Azzalini (1985, 1986) and Henze (1986). Generalizations to the multivariate case are given in Azzalini and Dalla-Valle (1996) and Azzalini and Capitanio (1999). Also, Arellano-Valle et al. (2002) show that many of the properties of the multivariate skew-normal distribution hold for a general class of skewed distributions. These are obtained from a symmetric class, defined in terms of independence conditions on signs and absolute values and give a general formula to obtain skewed pdf's. From these results, Arellano and Genton (2005) introduced the class of fundamental skewed distributions, and gave an unified approach to obtain multivariate skew distributions starting from symmetric ones. In this paper, we shall use a modified version of the multivariate skew-normal distribution, proposed by Azzalini and Dalla-Valle (1996) which is a special case of the fundamental skew-normal distribution proposed by Arellano-Valle and Genton (2005).

We assume a $p \times 1$ random vector $\boldsymbol{Y}$ follows a SN distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$ and $p \times 1$ skewness parameter vector $\boldsymbol{\lambda}$, and we write $\boldsymbol{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, if its probability density function (pdf) is given by

$$f(\mathbf{y}) = 2\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{3}$$

where $\phi_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the pdf of the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ shown as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Note that for $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = I_p$, expression (3) reduces to standardized multivariate skew-normal distribution with skewness parameter vector $\boldsymbol{\lambda}$, denoted by $SN_p(\boldsymbol{\lambda})$ also for $\boldsymbol{\lambda} = \mathbf{0}$ reduces to the symmetric $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-pdf, while for non-zero value of $\boldsymbol{\lambda}$, it produces a perturbed (asymmetric) family of $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-pdf's.

Expectation of $\boldsymbol{Y}$ if $\boldsymbol{Y} \sim SN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, is given by:

$$E[\boldsymbol{Y}] = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \times \boldsymbol{\delta}\sqrt{\frac{2}{\pi}}, \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}'\boldsymbol{\lambda}}},$$

(Arellano-Valle and Genton, 2005).

Some properties concerning this form of multivariate skew-normal distribution are referred by Azzalini and Della-Valle (1996), Azzalini and Capitanio (1999) and Arellano-Valle and Genton (2005).

Our focus in this article is on longitudinal data analysis. Longitudinal studies represent one of the principal research strategies employ in medical and social research. The defining feature of such studies is that subjects are measured repeatedly through time. A pervasive problem that arises in the context of analysis of longitudinal data is presence of missing data. In some cases, a subject may be missing one of several measurement occasions; however, it is more likely that there are missing data due to drop-out. Drop-out, refers to a subject removing himself or herself from the study, prior to the end of the study. Consequently, the data record for this subject prematurely terminates.

Rubin (1976) provided a framework for the incomplete data by introducing the important taxonomy of missing data mechanisms, consist of *missing completely at random* (MCAR), *missing at random* (MAR) and *missing not at random* (MNAR). A mechanism is said MCAR, if missing values are independent of both unobserved and observed data, and MAR if, conditional on the observed data, the missing values are independent of the missing measurements and otherwise the missing process is termed MNAR. In addition, Diggle and Kenward (1994) defined dropout process to be completely random dropout (CRD) if it is MCAR, random dropout (RD) if it is MAR and non-random dropout (NRD) if drop-out is dependent on missing outcomes.

Several simple approaches to this problem have been proposed, none of which are statistically satisfactory, for instance the simplest approach is *complete case analysis* where limits the analysis to only those subjects that are completed for all time points. Unfortunately, the available sample at the end of the study may have little resemblance to the sample initially randomized. Reasons for not completing the study may be confounded with the effects that the study was designed to investigate. Another method that belong to a class of semi-parametric regression techniques is *generalized estimating equations* (GEE). Method of GEE is used to fit the parameters of a generalized linear model where correlation between responses of the same individual has to be taken into account. The GEE allows for correlation without explicitly defining a model for the origin of the dependency, hence it is most suitable when the random effects and their variances are not of direct interest, also this method is useful under MCAR. They are frequently applied in longitudinal studies as they can handle many types of unmeasured dependence. In Section 3.2 these methods are explained and compared with proposed methodology in this paper.

Even though the assumption of likelihood ignorability encompasses both MAR and the more stringent and often implausible MCAR mechanisms, in most real settings it is impossible to exclude the possibility of a more general missingness mechanism, i.e. MNAR. A solution is to fit an MNAR model, consider a general pattern of missing data and let $\boldsymbol{R}_i$ denote the associated vector of missingness indicator related to $\boldsymbol{Y}_i$, such that $R_{ij} = 1$ if Y_{ij} is observed and otherwise $R_{ij} = 0$. When missing mechanism is MNAR, three frameworks modelling approaches may be used to joint model the missing mechanism and responses: selection, pattern-mixture and shared parameter models. These models are defined by the conditional factorization of the joint distribution of $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$.

The selection model factorization is based on

$$f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i|\boldsymbol{\theta})f(\mathbf{r}_i|\mathbf{y}_i, \boldsymbol{\psi}), \tag{4}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ denote parameter vector of measurements and missingness mechanism, respectively. The two vectors of parameters are assumed to be distinct. The first factor in (4) is the marginal density of the measurement process and the second one is the density of the missingness process, conditional on the outcomes.

An alternative model is based on so-called pattern-mixture models

(Little 1993, 1994, 1995). These are based on the factorization

$$f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i|\mathbf{r}_i,\ \boldsymbol{\theta})f(\mathbf{r}_i|\boldsymbol{\psi}).$$

The third model is referred to as shared-parameter models:

$$f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{\theta},\ \boldsymbol{\psi},\ \mathbf{b}_i) = f(\mathbf{y}_i|\mathbf{r}_i,\ \boldsymbol{\theta},\ \mathbf{b}_i)f(\mathbf{r}_i|\boldsymbol{\psi},\ \mathbf{b}_i),$$

where the model explicitly include a vector of unit-specific latent (or random) effects $\boldsymbol{b}_i$ of which one or more components are shared between both components in the joint distribution. Some references to such model are Wu and Carroll (1988) and Wu and Bailey (1988, 1989). Particular studies on non-random drop-out, when $\boldsymbol{Y}$ is distributed as normal distribution are considered by Diggle and Kenward (1994), Kenward (1998) and Crouchley and Ganjali (2002). In dropout pattern usually R_i is replaced by D_i, where D_i is the dropout location of the first missing value in Y_i.

Most of the above mentioned statistical methods are based on normal symmetric distribution and when asymmetric and consequently abnormality are observed in the data sets, some simple data-based transformations have proposed which maybe helpful in reducing skewness. Nevertheless the achievement of joint normality is rarely satisfied in multivariate cases. Furthermore, sometimes skewness is an inherent factor of population and ignoring them maybe lead to bias parameter estimates and subsequently misrepresentation of the results. Over the last two decades, there has been a growing interest in proposing the parametric family of multivariate skew-normal distribution that attracts the attention of researches to confront with skewness problems in the data sets. This distribution include multivariate normal family, when vector of skewness parameter is equal to zero, and have more flexibility of shape and some other characteristic such as skewness, kurtosis and dependence structure.

Some application of skew-normal family can be found in Arellano-Valle et al. (2005b) for measurement error model, Cancho et al. (2008) for non-linear regression model, Arellano-Valle et al. (2005a) and Lachos et al. (2007) for linear mixed model, Bolfarine et al. (2006) and Bolfarine and Lachos (2007) for binary regression and probit measurement error model, respectively. Also, there are some discussion about Baysian inference for skew-normal family in Liseo and Loperfido (2004).

Ibrahim et al. (2001) discussed nonignorable missingness in generalised linear mixed model. The use of skew-normal family for analyzing data with missing values has considered by Lin and Chen (2009) and

Baghfalaki and Ganjali (2011). Lin and Chen (2009) formulated Expectation Conditional Maximization (ECM) algorithm for calculating parameter estimation of incomplete skew data and they used Gibbs sampling for performing a Bayesian inference on model parameters to create multiple imputations for missing values. Baghfalaki and Ganjali (2011) developed EM algorithm to obtain the maximum likelihood estimates of parameters in bivariate skew-normal distribution for analyzing longitudinal skew-normal data with non-monotone missing values. However, both of these applications are under MAR mechanism. Although these methods will have reasonable results at the suitable condition, under MNAR (NRD) mechanism, they are unusable. In this paper, under a NRD assumption, we extend non-random dropout model of Diggle and Kenward (1994) for the analysis of multivariate skew-normal models when missing values occur in the responses.

In the context of non-random missingness the use of a selection model does not involve any identifiably problem for parameter estimation. However, as it is mentioned use of this method should be approached with caution (Glynn et al., 1986). The suggestion is the use of sensitivity analysis when one uses selection method (Molenberghs et al., 2003). We have some sensitivity analysis in the application Section.

The plan of this article is as follows, in Section 2, the selection model using multivariate skew-normal distribution is explained. Section 3 reports results of a simulation study, Section 4 reports applications to a real data sets indicating the usefulness of the approach and the last Section contains conclusions.

2 Non-ignorable Model for Longitudinal Data with Dropout: Using Multivariate Skew-Normal Distribution

The use of joint models for analyzing missing data in longitudinal studies first proposed by Heckman (1979) in the econometrics literature. More recently, Leigh et al. (1993) present an article on implementation of this approach. In its original formulation, the use of joint model involves two stages which are either performed separately. The first stage is to develop a predictive model for whether or not a subject drops out, using variables usually obtained prior to the dropout, often the variables measured at baseline or time-varying covariates. This model of dropout provides a predicted dropout probability or propensity score for each subject. These dropout propensity scores are then used in the second stage of longitudinal data model as a covariate to adjust for the potential influence of dropout. Stated in another way, this approach combines a marginal Gaussian regression model for the response, as might be used in the absence of missing data, with a Gaussian-based threshold model (in trust of a probit regression model) for the probability of a value being missing. This basic structure underlies the simplest form of a selection model that has been proposed for longitudinal data in the biometric setting by Diggle and Kenward (1994). Especially for a continuous response, these models can be constructed in a fairly obvious way, combining the multivariate Gaussian linear model with a suitable dropout model. Diggle and Kenward (1994) extend this approach by augmenting a logistic dropout model with past values of the dependent variable and multivariate Gaussian unobserved dependent variable at the time of dropout and this extent was also a development of the model of Greenlees et al. (1982) for nonrandom missingness in a cross-sectional setting.

Multivariate skew-normal distribution has good properties as mentioned in introduction (see, Azzalini, 1985, 1986, 2005). Furthermore in incomplete data literatures, because of missing values, one can not be insured about distribution of dependence variables. Since skew-normal family include normal one, estimation of skewness in skew-normal distribution is authorized to data and if data already have became from normal population, the skewness parameters will not be estimate significant, therefore using skew-normal distribution instead of normal one is more useful for missing responses in above models.

We consider a linear regression model where the error distribution follows the multivariate skew-normal distribution; also the probit regression model given outcomes is used as the probability of an observation to be missing. We obtain a compact form for the likelihood which does not need any approximation for integrals.

Suppose that we have n independent subjects, each corresponding to a vector $\boldsymbol{Y}_i$ of repeated measurements and each one be a sequence of measurements Y_{ij} designed to be measured at occasions j $(= 1, 2, ..., T)$. Let $\boldsymbol{Y}_i \sim SN_T(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. This experiment is a balanced experiment because of the same number of repeated measurements planned for all participating subjects, at fixed points in times. Associated with individual i we assume a known $T \times p$ covariate matrix $\boldsymbol{X}_i$, which we use to specify the linear predictor $\boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression coefficients. If dropout occurs, $\boldsymbol{Y}_i$ is only partially observed. We denote the occasion at which dropout occurs by $D_i > 1$ and $\boldsymbol{Y}_i$ is split into the $(D_i - 1)$-dimensional observed components, $\boldsymbol{Y}_i^o$, and the $(T - D_i + 1)$-dimensional missing components, $\boldsymbol{Y}_i^m$. In the case of no dropout, we let $D_i = T + 1$, and $\boldsymbol{Y}_i$ equals to $\boldsymbol{Y}_i^o$. In the selection model, the joint distribution of $\boldsymbol{Y}_i$ and D_i is factorized as the marginal distribution of $\boldsymbol{Y}_i$ and the conditional distribution of D_i given $\boldsymbol{Y}_i$, thus

$$f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|\boldsymbol{\theta})f(d_i|\boldsymbol{y}_i, \boldsymbol{\psi}),$$

then, the likelihood of the i^{th} subject based on the observed data $(\boldsymbol{y}_i^o, D_i = d_i)$, is proportional to the following marginal density function

$$\begin{aligned} f(\boldsymbol{y}_i^o,\ d_i|\boldsymbol{\theta},\ \boldsymbol{\psi}) &= \int f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta},\ \boldsymbol{\psi})d\boldsymbol{y}_i^m \\ &= \int f(\boldsymbol{y}_i|\boldsymbol{\theta}) \times f(d_i|\boldsymbol{y}_i,\ \boldsymbol{\psi})\ \boldsymbol{dy}_i^m. \end{aligned}$$

This terminology is due to Heckman (1979), which in that a marginal model for $\boldsymbol{Y}_i$ is combined with a model for the dropout process, conditional on the responses, and where $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are distinct vectors of unknown parameters in the measurement model and dropout model, respectively.

Let $\boldsymbol{h}_{ij} = (y_{i1}, ..., y_{i,j-1})$ denote the observed history of subject i up to time $t_{i,j-1}$, also suppose, the conditional probability for dropout at occasion j, given that the subject was still observed at previous occasion, to be dependent on the history $\boldsymbol{h}_{ij}$, through $y_{i,j-1}$, and the possibly unobserved current outcome y_{ij}, but not on the future outcomes $y_{ik}, k > j$. Therefore these conditional probabilities are given by:

$$p(D_i = j|D_i \geq j, \mathbf{h}_{ij}, y_{ij}, \boldsymbol{\psi}) = \Phi(\psi_0 + \psi_1 y_{ij} + \psi_2 y_{i,j-1}). \qquad (5)$$

The special case of model (5) corresponding to RD and CRD are obtained from setting $\psi_1 = 0$ and $\psi_1 = \psi_2 = 0$, respectively. In the first case, dropout is no longer allowed to be dependent on the current measurement, and in the second case, dropout is independent of all outcomes. This probability depends on the last observed response and the current value of response which if not missing, could have been observed. The ψ_1 and ψ_2 are parameters which should be estimated.

The main result of this paper is based on the following proposition by which we obtain the observed likelihood function, $f(\boldsymbol{y}^o, d|\boldsymbol{\theta},\ \boldsymbol{\psi})$. Here

$\boldsymbol{Y}$ is distributed as multivariate skew-normal with location parameter $\boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta}$, dispersion matrix $\boldsymbol{\Sigma}$ and skewness parameter vector $\boldsymbol{\lambda}$.

The following lemma, proved by Arellano-Valle et al. (2005a), is a useful lemma that will be used in the following proposition.

Lemma 2.1. *Let $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then, for any fixed k-dimensional vector $\mathbf{a}$ and $k \times n$ matrix $\boldsymbol{B}$,*

$$E[\Phi_k(\mathbf{a} + \mathbf{BY}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = \Phi_k(\mathbf{a}|\boldsymbol{\eta} - \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B\Sigma B}').$$

Proposition 2.1. *Let $\boldsymbol{Y}_i \sim SN_T(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, $\boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta}$, and we have model (5) for dropout. Then the joint distribution of the observed measurement and missing process for $j = 3, ..., T$ (see Molenberghs and Kenward, 2007, pp. 186) is given by:*

$$\begin{aligned} f(\boldsymbol{y}_i^o, d_i = j|\boldsymbol{\theta}, \boldsymbol{\psi}) &= \left(\prod_{k=2}^{j-1}[1 - \Phi(\psi_0 + \psi_1 y_{ik} + \psi_2 y_{i(k-1)})]\right) \\ &\times \ (2\phi_{j-1}(\boldsymbol{y}_i^o|\boldsymbol{\mu}_{i1},\ \boldsymbol{\Sigma}_{11}) \\ &\times \ \Phi_2(\mathbf{A}_{ij} + \boldsymbol{u}\boldsymbol{\mu}_{i2}^c|\mathbf{0}, \boldsymbol{I} + \mathbf{u}\boldsymbol{\Sigma}_{22.1}\mathbf{u}')) \end{aligned}$$

where, if $\boldsymbol{\zeta} = (\psi_1, 0, 0, \cdots, 0)$ is a vector of dimension $T - j + 1$, then

$$\boldsymbol{\gamma}' = (\boldsymbol{\gamma}_1' \ \boldsymbol{\gamma}_2') = \boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1/2}, \quad \mathbf{A}_{ij} = \begin{pmatrix} \psi_0 + \psi_1 y_{i,j-1} \\ \boldsymbol{\gamma}_1'\boldsymbol{y}_i^o - \boldsymbol{\gamma}'\boldsymbol{\mu}_i \end{pmatrix}, \quad \boldsymbol{u} = \begin{pmatrix} \boldsymbol{\zeta}' \\ \boldsymbol{\gamma}_2' \end{pmatrix}.$$

and

$$\begin{aligned} \boldsymbol{\mu}_{i2}^c &= \boldsymbol{\mu}_{i2} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_{i1}), \\ \boldsymbol{\Sigma}_{22.1} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}. \end{aligned}$$

where $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{i1}' \ \ \boldsymbol{\mu}_{i2}')'$ is the partitioned location vector of two sets of variables, $\boldsymbol{Y}_i^o$ and $\boldsymbol{Y}_i^m$, where $\boldsymbol{\mu}_{i1}$ is the location vector for variables in $\boldsymbol{Y}_i^o$ and $\boldsymbol{\mu}_{i2}$ is the location vector for variables in $\boldsymbol{Y}_i^m$, also

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

is the partitioned scale matrix for these variables, where $\boldsymbol{\Sigma}_{11}$ is $(j-1) \times (j-1)$, corresponds to variables $\boldsymbol{Y}_i^o$, $\boldsymbol{\Sigma}_{22}$ is $(T - j + 1) \times (T - j + 1)$, *corresponds to* $\boldsymbol{Y}_i^m$*, and* $\boldsymbol{\Sigma}_{12}$ *is* $(j-1)\times(T-j+1)$*, corresponds to scale matrix of* $\boldsymbol{Y}_i^o$ *and* $\boldsymbol{Y}_i^m$*.*

Proof. Suppose $\boldsymbol{y}_i^o = (y_{i1}, y_{i2}, ..., y_{i(j-1)})'$ and $\boldsymbol{y}_i^m = (y_{ij}, y_{i(j+1)}, ..., y_{iT})'$, we have

$$
\begin{aligned}
& f(\boldsymbol{y}_i^o, d_i = j|\boldsymbol{\theta}, \boldsymbol{\psi}) \\
= & \int \left(\Phi(\psi_0 + \psi_1 y_{ij} + \psi_2 y_{i,j-1}) \times \prod_{k=2}^{j-1} [1 - \Phi(\psi_0 + \psi_1 y_{ik} + \psi_2 y_{i(k-1)})] \right) \\
\times & \; 2\phi_T(\boldsymbol{y}_i|\boldsymbol{\mu}_i,\ \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_i))d\boldsymbol{y}_i^m \\
= & \left(\prod_{k=2}^{j-1} [1 - \Phi(\psi_0 + \psi_1 y_{ik} + \psi_2 y_{i(k-1)})] \right) \times 2\phi_{j-1}(\boldsymbol{y}_i^o|\boldsymbol{\mu}_{i1},\ \boldsymbol{\Sigma}_{11}) \\
\times & \int \Phi_2(\mathbf{A}_{ij} + \mathbf{u}\boldsymbol{y}_i^m|\mathbf{0}, \boldsymbol{I})\phi_{T-j+1}(\boldsymbol{y}_i^m|\boldsymbol{\mu}_{i2}^c, \boldsymbol{\Sigma}_{22.1})d\boldsymbol{y}_i^m \\
= & \left(\prod_{k=2}^{j-1} [1 - \Phi(\psi_0 + \psi_1 y_{ik} + \psi_2 y_{i(k-1)})] \right) \\
\times & \; 2\phi_{j-1}(\boldsymbol{y}_i^o|\boldsymbol{\mu}_{i1},\ \boldsymbol{\Sigma}_{11}) \times \Phi_2(\mathbf{A}_{ij} + \boldsymbol{u}\boldsymbol{\mu}_{i2}^c|\mathbf{0}, \boldsymbol{I} + \mathbf{u}\boldsymbol{\Sigma}_{22.1}\mathbf{u}'),
\end{aligned}
$$

since,

$$
\begin{aligned}
& \Phi(\psi_0 + \psi_1 y_{ij} + \psi_2 y_{i,j-1}) \times \Phi(\boldsymbol{\gamma}_1'\boldsymbol{y}_i^o + \boldsymbol{\gamma}_2'\boldsymbol{y}_i^m - \boldsymbol{\gamma}'\boldsymbol{\mu}_i) \\
&= \Phi_2\left(\begin{pmatrix} \psi_0 + \psi_1 y_{ij} + \psi_2 y_{i,j-1} \\ \boldsymbol{\gamma}_1'\boldsymbol{y}_i^o + \boldsymbol{\gamma}_2'\boldsymbol{y}_i^m - \boldsymbol{\gamma}'\boldsymbol{\mu}_i \end{pmatrix} |\mathbf{0}, \boldsymbol{I} \right) \\
&= \Phi_2\left(\begin{pmatrix} \psi_0 + \psi_2 y_{i,j-1} \\ \boldsymbol{\gamma}_1'\boldsymbol{y}_i^o - \boldsymbol{\gamma}'\boldsymbol{\mu}_i \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}' \\ \gamma_2' \end{pmatrix} \boldsymbol{y}_i^m |\mathbf{0}, \boldsymbol{I} \right) \\
&= \Phi_2\left(\mathbf{A}_{ij} + \mathbf{u}\boldsymbol{y}_i^m|\mathbf{0}, \boldsymbol{I}\right).
\end{aligned}
$$

In selection models, inference is based on the joint distribution of missingness indicator and observed outcome. Therefore, the corresponding likelihood function can be written as multiplication of the following observed density function for all individual:

$$f(\boldsymbol{y}_i^o, d_i = j|\boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} 2\phi_{j-1}(\boldsymbol{y}_i^o|\boldsymbol{\mu}_{i1}, \boldsymbol{\Sigma}_{11}) & j = 2 \\ \quad \times\Phi_2(\boldsymbol{A}_{ij} + \boldsymbol{u}\boldsymbol{\mu}_{i2}^c|\boldsymbol{0}, \boldsymbol{I} + \boldsymbol{u}\boldsymbol{\Sigma}_{22.1}\boldsymbol{u}') & \\ & \\ 2\phi_{j-1}(\boldsymbol{y}_i^o|\boldsymbol{\mu}_{i1}, \boldsymbol{\Sigma}_{11}) & j = 3, ..., T \\ \quad \times\Phi_2(\boldsymbol{A}_{ij} + \boldsymbol{u}\boldsymbol{\mu}_{i2}^c|0, I + \boldsymbol{u}\boldsymbol{\Sigma}_{22.1}\boldsymbol{u}') & \\ \quad \times \prod_{k=2}^{j-1}[1 - \Phi(\psi_0 + \boldsymbol{\psi}_1 y_{ik} + \boldsymbol{\psi}_2 y_{i(k-1)})] & \\ & \\ 2\phi_T(\boldsymbol{y}_i|\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}) & j = T + 1 \\ \quad \times\Phi(\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})) & \\ \quad \times \prod_{k=2}^{T}[1 - \Phi(\psi_0 + \boldsymbol{\psi}_1 y_{ik} + \boldsymbol{\psi}_2 y_{i(k-1)})] & \end{cases}$$

Computation of the derivation of $\ell = ln(L_{obs})$, needed for optimization and for estimation of parameters may be followed by the same method of Diggle and Kenward (1994). This likelihood can be maximized directly by using existing statistical softwares such as R or S-plus or Matlab or other programmable software, for example in R and Matlab, one may use function nlminb and fmincon, respectively, for optimization of observed likelihood.

3 Simulation Study

3.1 Simulation Study for Joint Model

In this section, a simulation study was conducted to evaluate the usefulness of the proposed model. The study was replicated three times: first assuming completely random dropout, then assuming random dropout and finally assuming non-random dropout. The criteria used for comparison is the standardized bias of the estimators which introduced by Demirtas (2007), according to him: If the parameter of interest is θ, the standardized bias is $\frac{E(\hat{\theta})-\theta}{SE(\hat{\theta})}$, where SE stands for standard error. If the standardized bias exceeds 0.4 in a positive or negative direction, then the bias begins to have a noticeable adverse impact on efficiency. Let start the studies.

1. Completely Random Dropout Assumption
   In our simulation study we generate 500 samples with sample size $n = 100$ and $n = 300$. Here, we have simulated data according to

the following longitudinal linear regression model with two time point:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 u_{ij} + \beta_3 t_{ij} u_{ij} + \varepsilon_{ij}, \quad i : 1, 2, ..., n, \quad j : 1, 2$$

where $t_{ij} = j$, were coded $1, 2$ for two time points and u_{ij} were uniformly generated from -3 to 3. The regression coefficients were defined to be: $\beta_0 = 5$, $\beta_1 = -2$, $\beta_2 = 0$, $\beta_3 = 1$, and the following scale and skewness parameters

$$\boldsymbol{\Sigma}^{1/2} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

also, we consider three rates of missingness 10%, 20% and 30%. For this purpose, we consider the following ignorable dropout mechanisms

$$p(D = 2|y_1, y_2, \boldsymbol{\psi}_k) = \Phi(\psi_{0k}), \quad k = 1, 2, 3. \tag{6}$$

Such that, for $\psi_{01} = -1.281$, $\psi_{02} = -0.842$ and $\psi_{03} = -0.524$ rates of missing are 0.1, 0.2 and 0.3, respectively. The parameter estimates, standard errors and standard biases of this simulation are presented in Table 1. All of the results confirm goodness of fit of the proposed methodology. Note that in all tables of this section D_{11}, D_{12} and D_{22} are the distinct elements of the matrix $\boldsymbol{\Sigma}^{1/2}$. There is no considerable bias in Table 1.

2. Random Dropout Assumption
Simulation study for this part is similar to simulation under CRD assumption in previous part with one difference that the following ignorable mechanisms is used instead of equation (6),

$$p(D = 2|y_1, y_2, \boldsymbol{\psi}_k) = \Phi(\psi_{0k} + \psi_{2k} y_1), \quad k = 1, 2, 3. \tag{7}$$

Such that, for $\psi_{01} = 1$ and $\psi_{21} = -0.95$ expected rate of missingness is 0.10, for $\psi_{02} = 1$ and $\psi_{22} = -0.7$ expected rate of missingness is 0.20 and finally for $\psi_{03} = 1$ and $\psi_{23} = -0.4$ expected missing response is 0.30. Table 2 shows the results of this simulation. There is no considerable bias for this case.

3. Non-Random Dropout Assumption
This simulation study is under NRD and is similar to the pervious parts. The following non-ignorable mechanism are considered

instead of equation (6) or (7),

$$p(D = 2|y_1, y_2, \boldsymbol{\psi}_k) = \Phi(\psi_{0k} + \psi_{1k}y_2 + \psi_{2k}y_1), \quad k = 1, 2, 3. \quad (8)$$

also for three rate of missingness, 0.10, 0.20 and 0.30, the sets of parameters $\psi_{01} = 1$, $\psi_{11} = -0.3$ and $\psi_{21} = -0.9$; $\psi_{02} = 1$, $\psi_{12} = -0.7$ and $\psi_{22} = -0.6$; $\psi_{03} = 1$, $\psi_{13} = -0.4$ and $\psi_{23} = -0.4$ are used, respectively. Results of this study are given in Table 3. Considerable bias is for one element of $\boldsymbol{\Sigma}^{1/2}$ (D_{12}) which happens for moderate sample size ($n = 100$) with high missing rate.

3.2 Generalized Estimating Equation Simulation

One viewpoint, in achievement a parametric class of distribution, is increasing the degree of flexibility by suitability increasing the number of parameters. Such a construction, under appropriate conditions on the achievable degree of approximation to an arbitrary target distribution, builds a bridge between the parametric and the non-parametric context and the skew-normal distribution is the useful one that adapted to this concept (Azzalini, 2005). This outline prepare the way for comparison between our methodology in employing multivariate skew-normal distribution for analyzing incomplete data via the joint model and a famous class of semi-parametric method such as generalized estimating equation (GEE).

The analysis of correlated data arising from repeated measurements when the measurements are assumed to be multivariate normal has been studied extensively. However, the normality assumption might not always be reasonable. GEE provide a practical method for estimating regression and association parameters without specify the entire likelihood to analyze such data.

Liang and Zeger (1986) introduced GEE as a method of dealing with correlated data when, except for the correlation among responses, the data can be modelled as a generalized linear model. The GEE methodology is based on solving the equations

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} (V(\boldsymbol{y}_i))^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \quad (9)$$

where, the marginal covariance matrix $V(\boldsymbol{y}_i)$ is decomposed into $\phi A_i^{1/2} R_i A_i^{1/2}$, with A_i the matrix with the marginal variances on the main diagonal and zeros elsewhere, $R_i = R_i(\boldsymbol{\alpha})$ the marginal correlation matrix,

Table 1: Results of Simulation Study for 500 Samples under Complectly Random Dropout, (Bold numbers are estimated parameters with standardized bias > 0.4).

Model		Parametricmodel						GEE					
rateof missing		%10		%20		%30		%10		%20		%30	
	True	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std. B.	Est.(S.E.)	Std.B.
Parameter	Samplesize $n = 100$												
β_0	5.000	4.994(0.046)	-0.123	4.994(0.054)	-0.105	4.981(0.066)	-0.287	5.615(0.187)	**3.297**	5.613(0.193)	**3.181**	5.637(0.198)	**3.215**
β_1	-2.000	-1.998(0.045)	0.026	-1.997(0.047)	0.046	-2.006(0.080)	-0.075	-1.987(0.112)	0.117	-1.989(0.118)	0.095	-2.006(0.123)	-0.051
β_2	0.000	-0.001(0.039)	-0.025	-0.004(0.049)	-0.070	0.002(0.077)	0.029	-0.015(0.116)	-0.128	0.003(0.116)	0.025	0.007(0.110)	0.065
β_3	1.000	0.992(0.033)	-0.226	0.986(0.044)	-0.308	0.973(0.071)	-0.367	1.006(0.076)	0.078	0.997(0.078)	-0.035	0.996(0.069)	-0.055
D_{11}	1.000	0.993(0.041)	-0.170	0.983(0.051)	-0.330	0.983(0.071)	-0.233	-	-	-	-	-	-
D_{12}	0.400	0.395(0.038)	-0.113	0.388(0.045)	-0.251	0.383(0.077)	-0.219	-	-	-	-	-	-
D_{22}	1.000	1.009(0.042)	0.234	1.014(0.055)	0.255	1.034(0.084)	**0.409**	-	-	-	-	-	-
λ_1	2.000	2.002(0.046)	0.054	1.999(0.049)	-0.018	2.002(0.083)	0.026	-	-	-	-	-	-
λ_2	2.000	1.986(0.045)	-0.294	1.991(0.049)	-0.184	1.977(0.086)	-0.263	-	-	-	-	-	-
Parameter	Samplesize $n = 300$												
β_0	5.000	4.996(0.035)	-0.126	4.995(0.053)	-0.089	4.964(0.21)	-0.167	5.633(0.105)	**6.009**	5.628(0.117)	**5.353**	5.648(0.106)	**6.135**
β_1	-2.000	-2.001(0.031)	-0.023	-1.998(0.047)	0.041	-1.996(0.130)	0.026	-2.002(0.067)	-0.033	-1.998(0.074)	0.017	-2.012(0.068)	-0.183
β_2	0.000	0.001(0.027)	0.022	-0.000(0.042)	-0.011	-0.020(0.215)	-0.093	0.009(0.062)	0.161	-0.011(0.064)	-0.167	-0.006(0.076)	-0.075
β_3	1.000	0.991(0.023)	-0.385	0.984(0.048)	-0.328	0.991(0.214)	-0.042	0.994(0.039)	-0.145	1.005(0.040)	0.115	1.003(0.048)	0.052
D_{11}	1.000	0.995(0.029)	-0.167	0.985(0.046)	-0.326	0.978(0.210)	-0.101	-	-	-	-	-	-
D_{12}	0.400	0.396(0.027)	-0.130	0.388(0.045)	-0.256	0.366(0.209)	-0.163	-	-	-	-	-	-
D_{22}	1.000	1.009(0.033)	0.273	1.013(0.048)	0.268	1.077(0.216)	0.358	-	-	-	-	-	-
λ_1	2.000	2.001(0.034)	0.017	1.995(0.047)	-0.108	2.012(0.223)	0.054	-	-	-	-	-	-
λ_2	2.000	1.994(0.038)	-0.163	1.984(0.054)	-0.287	1.962(0.220)	-0.171	-	-	-	-	-	-

Table 2: Results of Simulation Study for 500 Samples under Random Dropout, (Bold numbers are estimated parameters with standardized bias > 0.4).

Model	Parametricmodel							GEE					
rateof missing		%10		%20		%30		%10		%20		%30	
	True	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std. B.	Est.(S.E.)	Std.B.
Parameter	Samplesize$n$ = 100												
β_0	5.000	5.004(0.044)	0.083	4.987(0.069)	-0.164	4.995(0.065)	-0.074	5.859(0.192)	**4.478**	5.801(0.113)	**7.072**	6.132(0.195)	**5.809**
β_1	-2.000	-2.004(0.045)	-0.092	-1.991(0.070)	0.116	-2.002(0.070)	-0.032	-2.115(0.121)	**−0.945**	-2.086(0.071)	**−1.215**	-2.255(0.129)	**−1.967**
β_2	0.000	0.002(0.039)	0.043	0.009(0.063)	0.140	-0.009(0.062)	-0.147	-0.184(0.109)	**−1.694**	-0.045(0.069)	**−0.644**	-0.292(0.141)	**−2.074**
β_3	1.000	0.989(0.034)	-0.322	0.979(0.058)	-0.350	0.983(0.046)	-0.370	1.095(0.064)	**1.469**	1.022(0.046)	**0.481**	1.145(0.084)	**1.725**
D_{11}	1.000	0.988(0.041)	-0.278	0.989(0.058)	-0.179	0.982(0.061)	-0.293	-	-	-	-	-	-
D_{12}	0.400	0.392(0.038)	-0.200	0.391(0.046)	-0.183	0.383(0.057)	-0.291	-	-	-	-	-	-
D_{22}	1.000	1.007(0.051)	0.135	1.015(0.063)	0.240	1.019(0.062)	0.313	-	-	-	-	-	-
λ_1	2.000	2.003(0.055)	0.067	2.004(0.063)	0.062	1.999(0.078)	-0.013	-	-	-	-	-	-
λ_2	2.000	2.001(0.045)	0.026	1.993(0.071)	-0.101	1.989(0.076)	-0.139	-	-	-	-	-	-
Parameter	Samplesize$n$ = 300												
β_0	5.000	4.997(0.033)	-0.086	5.006(0.041)	0.167	5.011(0.087)	0.127	5.871(0.102)	**8.520**	5.780(0.197)	**3.953**	6.116(0.129)	**8.667**
β_1	-2.000	-1.999(0.031)	0.009	-1.976(0.063)	0.380	-1.998(0.072)	0.026	-2.126(0.064)	**−1.981**	-2.082(0.127)	**−0.647**	-2.241(0.082)	**−2.933**
β_2	0.000	0.006(0.030)	0.206	0.011(0.038)	0.304	0.009(0.065)	1.137	-0.175(0.068)	**−2.572**	-0.046(0.126)	-0.369	-0.302(0.083)	**−3.638**
β_3	1.000	0.989(0.028)	-0.386	0.987(0.058)	-0.214	0.976(0.069)	-0.336	1.088(0.042)	**2.093**	1.023(0.075)	0.311	1.154(0.050)	**3.064**
D_{11}	1.000	0.994(0.028)	-0.192	0.991(0.038)	-0.129	0.983(0.061)	-0.266	-	-	-	-	-	-
D_{12}	0.400	0.397(0.029)	-0.090	0.395(0.031)	-0.145	0.379(0.065)	-0.321	-	-	-	-	-	-
D_{22}	1.000	1.009(0.034)	0.270	0.995(0.038)	-0.129	1.016(0.081)	0.199	-	-	-	-	-	-
λ_1	2.000	1.995(0.029)	-0.163	1.991(0.042)	-0.200	1.989(0.082)	-0.133	-	-	-	-	-	-
λ_2	2.000	1.995(0.033)	-0.141	1.994(0.034)	-0.178	1.983(0.088)	-0.186	-	-	-	-	-	-

Table 3: Results of Simulation Study for 500 Samples under Random Dropout, (Bold numbers are estimated parameters with standardized bias > 0.4).

Model	Parametricmodel							GEE					
rateof missing		%10		%20		%30		%10		%20		%30	
	True	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std.B.	Est.(S.E.)	Std. B.	Est.(S.E.)	Std.B.
Parameter	Samplesize $n = 100$												
β_0	5.000	4.989(0.084)	-0.126	5.002(0.157)	0.012	5.014(0.157)	0.088	5.574(0.185)	**3.105**	5.660(0.193)	**3.410**	5.359(0.291)	**1.234**
β_1	-2.000	-1.972(0.070)	0.399	-1.964(0.130)	0.277	-1.967(0.131)	0.253	-1.940(0.116)	**0.517**	-1.971(0.127)	0.224	-1.711(0.216)	**1.340**
β_2	0.000	-0.008(0.068)	-0.123	0.001(0.125)	0.007	-0.018(0.123)	-0.152	0.063(0.108)	**0.585**	0.043(0.127)	0.337	0.153(0.169)	**0.905**
β_3	1.000	0.988(0.049)	-0.240	0.979(0.096)	-0.212	0.981(0.103)	-0.182	0.940(0.069)	**−0.860**	0.949(0.084)	**−0.605**	0.842(0.128)	**−1.234**
D_{11}	1.000	0.984(0.065)	-0.238	0.982(0.113)	-0.159	0.993(0.121)	-0.052	-	-	-	-	-	-
D_{12}	0.400	0.375(0.067)	-0.375	0.330(0.115)	**−0.607**	0.334(0.111)	**−0.600**	-	-	-	-	-	-
D_{22}	1.000	0.996(0.073)	-0.056	1.005(0.141)	0.038	0.988(0.142)	-0.081	-	-	-	-	-	-
λ_1	2.000	1.986(0.085)	-0.168	1.946(0.163)	-0.325	1.992(0.207)	-0.040	-	-	-	-	-	-
λ_2	2.000	1.979(0.091)	-0.233	1.998(0.166)	-0.007	1.962(0.159)	-0.241	-	-	-	-	-	-
Parameter	Samplesize $n = 300$												
β_0	5.000	5.002(0.027)	0.080	5.001(0.072)	0.013	4.978(0.179)	-0.124	5.562(0.110)	**5.095**	5.631(0.114)	**5.524**	5.316(0.162)	**1.946**
β_1	-2.000	-1.993(0.023)	0.291	-1.966(0.075)	0.453	-1.933(0.142)	0.472	-1.929(0.070)	**1.001**	-1.951(0.073)	**0.662**	-1.684(0.118)	**2.669**
β_2	0.000	-0.001(0.022)	-0.045	0.000(0.064)	0.000	-0.002(0.142)	-0.016	0.061(0.069)	**0.878**	0.040(0.069)	**0.576**	0.164(0.088)	**1.866**
β_3	1.000	0.994(0.018)	-0.309	0.999(0.076)	-0.013	0.965(0.116)	-0.298	0.939(0.044)	**−1.371**	0.950(0.047)	**−1.050**	0.835(0.067)	**−2.464**
D_{11}	1.000	1.000(0.023)	-0.009	0.998(0.079)	-0.025	0.982(0.121)	-0.144	-	-	-	-	-	-
D_{12}	0.400	0.397(0.019)	-0.174	0.371(0.091)	-0.318	0.355(0.111)	-0.396	-	-	-	-	-	-
D_{22}	1.000	0.997(0.023)	-0.114	0.996(0.072)	-0.055	1.001(0.145)	0.008	-	-	-	-	-	-
λ_1	2.000	1.997(0.024)	-0.135	1.973(0.084)	-0.321	1.954(0.167)	-0.273	-	-	-	-	-	-
λ_2	2.000	1.994(0.026)	-0.244	1.998(0.087)	-0.022	1.944(0.179)	-0.314	-	-	-	-	-	-

often referred to as the working correlation matrix, and ϕ an overdispersion parameter. Some discussion of GEE method can be found in Hardin and Hilbe (2003), Molenberghs and Kenward (2007) and Diggle et al. (2002). These GEE define an unbiased estimator only under MCAR (CRD).

We also conduct a simulation study using GEE for estimation of regression coefficient in the skew-normal longitudinal regression model. For this purpose, we have considered a model and real values the same as those in section 3.1 and we use the standard bias as a criterian to compare operation of joint model methodology and GEE for bivariate skew-normal longitudinal regression model. The Six last columns of tables 1-3 show the results of this simulation study. These results can be computed using SAS, proc genmod, or MATLAB, GEEQBOX toolbox. Results show, in spite of the relatively good operation of GEE in CRD, that the parameters estimate are bias in RD and NRD situation. Thus using joint model is more efficient than that of GEE for skew population. In addition, the joint model can estimate scale parameter, skewness parameter, and can recognize missingness mechanism in the data set.

In this section, we have simulated from GEE but one can compare results of GEE with other semi-parametric method such as WGEE or inverse probability weighted Generalized Estimating Equations WGEE (Robins et al., 1995; Yi and Cook, 2002) for future works.

4 Applications

4.1 Application 1: Mastitis in Dairy Cattle

We illustrate the usefulness of the proposed method by applying it to Mastitis data. These data, concerning the occurrence of the infectious disease called Mastitis in dairy cows, was introduced by Diggle and Kenward (1994). Data were available of the milk yields (in thousands of litres) of 107 dairy cows from a single herd in two consecutive years. In the first year all animals were free of Mastitis, in the second year 27 became infected. Mastitis typically reduces milk yield and these are considered as missing data. Further background details are given in Diggle and Kenward (1994) and in its accompanying discussion. An additional covariate, the year of first lactation of each cow, is ignored here as its inclusion in the analysis had a negligible impact (see Kenward, 1998, Crouchley and Ganjali, 2002). Clearly the occurrence of infection in second year can be represented as one of dropout in longitudinal data, with the occurrence of Mastitis corresponding to dropout. In addition, Molenberghs et al. (2001) and Crouchley and Ganjali (2002) found 3 outliers (cows 4, 5 and 66) in these data.

Using bivariate skew-normal distribution, we consider two situations. Firstly whole data are considered where using likelihood ratio test we found a CRD mechanism. Parameter estimates and standard errors under three different models (NRD, RD and CRD) are given in Table 6. We use observed information matrix for calculating standard errors. According to the expectation of multivariate skew-normal distribution, in Mastitis data by using the results of Table 6, $\hat{E}[\boldsymbol{Y}] = (5.762, 6.606)'$.

Secondly, for our subsequent analysis, we deleted outliers, results of this situation (under NRD, RD and CRD) are presented in Table 7. Also, for these data a CRD mechanism is found and $\hat{E}[\boldsymbol{Y}] = (5.798, 6.724)'$.

Note that, using bivariate skew-normal distribution a CRD mechanism is obtained. These results are interesting because, before this for Mastitis data a NRD mechanism is found (Kenward, 1998; Diggle and Kenward, 1994; Ganjali and Ranji, 2008) using bivariate normal distribution. Tables 8 and 9 show the results of normal model. These different results are a consequence of using bivariate skew-normal distribution and the existence of a significant skewness parameter. Notice also that Kenward (1998), using a t distribution for the response in the second year, find a RD mechanism for whole data.

Figure 1 displays the scatter plot of (Y_1, Y_2) for whole data with superimposed contours of the fitted skew-normal (right panel) and normal distribution (left panel) after fitting the joint models (the results of Table 6 and Table 8), which confirm similar fitness in normal and skew-normal model for the whole Mastitis data. Figure 2 displays the same plot as that of Figure 1 for the results of mastitis data after removing outliers.

There are some important points which have to be mentioned: at first, about the estimated variance of the parameter estimates in using normal and skew-normal distribution, we would like to mention that the increase of standard errors of using skew-normal distribution may be due to estimating two more parameters related to skewness. Also, consider that the roles of parameters in using normal and skew-normal distribution are different. For example μ_1 is the mean of Y_1 in using normal distribution, but this parameter is not the mean of Y_1 in using skew-normal distribution. Tables 6 and 8, also, show the results using RD and CRD assumptions. As it can be seen the results of the skew-normal distribution gives a better fit to the data (deviance= 622.718-

Table 4: Parameter estimates and standard errors under CRD, RD and NRD assumptions for analyzing the whole Mastitis data using bivariate skew normal distribution for response, σ_{11}, σ_{12} and σ_{22} are distinct components of $\Sigma^{1/2}$.

	NRD		RD		CRD	
Parameter	Est.	S. E.	Est.	S. E.	Est.	S. E.
μ_1	5.955	0.270	5.973	0.268	5.974	0.269
μ_2	5.789	0.307	5.763	0.305	5.763	0.305
σ_{11}	0.927	0.093	0.931	0.096	0.931	0.096
σ_{12}	0.204	0.092	0.207	0.084	0.207	0.084
σ_{22}	1.402	0.294	1.330	0.193	1.330	0.193
λ_1	-1.021	0.540	-1.023	0.543	-1.023	0.544
λ_2	1.860	0.786	1.724	0.657	1.724	0.658
ψ_0	-2.300	1.714	-1.603	0.860	-0.667	0.132
ψ_1	0.272	0.507	-	-	-	-
ψ_2	-0.033	0.376	0.161	0.145	-	-
$-2logL$	614.885		615.107		616.349	

615.107=7.611, with two degrees of freedom, p-value=0.022) under RD assumption. Results under assumption of CRD also show a better fit of using skew-normal distribution. The main point is that the use of normal distribution (which cannot find out the skewness of the data) misleads one to find that missing mechanism is not random. We notice that the use of the correct distribution (skew-normal distribution) gives a CRD mechanism. Note that skew-normal family include normal family so if the nature of data be symmetric, that reflect with nonsignificant skewness parameter. This does not occur in our data set. For one who uses normal distribution effect of skewness goes to a significant nonignorable parameter in dropout mechanism which gives a misleading conclusion (although gives nearly the same -2log likelihood).

Finally we investigate the sensitivity of the parameter estimates of μ_2 (see Molenberghs et al., 2001, for the importance of this) and value of $-2logL$ of the model by imposing different values for non-ignitability parameter of ψ_1 in model (5). For this we fix values ψ_1 for $\psi_1 : -0.5, -0.045, ..., 0.5$ and found estimate of μ_2 and $-2logL$; these results, presented in Table 10, confirm a RD mechanism for Mastitis data as $-2logL$ and the estimate of μ_2 do not change considerably.

Table 5: Parameter estimates and standard errors under CRD, DR and NRD assumptions for analyzing the Mastitis data without outliers using bivariate skew normal distribution for response, σ_{11}, σ_{12} and σ_{22} are distinct components of $\Sigma^{1/2}$.

	NRD		RD		CRD	
Parameter	Est.	S. E.	Est.	S. E.	Est.	S. E.
μ_1	5.168	0.212	5.145	0.197	5.145	0.197
μ_2	6.571	0.442	6.459	0.272	6.459	0.272
σ_{11}	1.029	0.135	1.046	0.131	1.046	0.131
σ_{12}	0.312	0.090	0.303	0.095	0.303	0.095
σ_{22}	1.027	0.143	0.999	0.087	0.999	0.087
λ_1	2.634	1.141	2.616	1.116	2.616	1.116
λ_2	-1.066	0.819	-1.009	0.734	-1.009	0.734
ψ_0	-2.103	1.743	-1.497	0.908	-0.644	0.132
ψ_1	0.457	1.210	-	-	-	-
ψ_2	-0.268	1.166	0.146	0.153	-	-
$-2logL$	546.382		546.518		547.429	

Table 6: Parameter estimates and standard errors under CRD, DR and NRD assumptions and bivariate normal model for analyzing the whole Mastitis Data, σ_{11}, σ_{12} and σ_{22} are distinct components of $\Sigma^{1/2}$.

	NRD		RD		CRD	
Parameter	Est.	S. E.	Est.	S. E.	Est.	S. E.
μ_1	5.765	0.090	5.765	0.090	5.765	0.090
μ_2	6.080	0.146	6.484	0.122	6.484	0.121
σ_{11}	0.894	0.060	0.877	0.056	0.877	0.058
σ_{12}	0.260	0.056	0.312	0.053	0.313	0.053
σ_{22}	1.246	0.113	1.094	0.083	1.094	0.083
ψ_0	0.369	1.433	-1.603	0.860	-0.667	0.131
ψ_1	-1.499	0.440	-	-	-	-
ψ_2	1.298	0.392	0.161	0.146	-	-
$-2logL$	617.490		622.718		623.962	

Table 7: Parameter estimates and standard errors under CRD, DR and NRD assumptions and bivariate normal nodel for analyzing the Mastitis data without outliers, σ_{11}, σ_{12} and σ_{22} are distinct components of $\Sigma^{1/2}$.

	NRD		RD		CRD	
Parameter	Est.	S. E.	Est.	S. E.	Est.	S. E.
μ_1	5.798	0.085	5.798	0.085	5.7985	0.085
μ_2	6.415	0.445	6.399	0.110	6.3993	0.110
σ_{11}	0.786	0.053	0.787	0.052	0.7873	0.052
σ_{12}	0.376	0.053	0.375	0.049	0.3752	0.049
σ_{22}	0.973	0.093	0.971	0.073	0.9719	0.073
ψ_0	-1.599	2.924	-1.497	0.908	0.6445	0.132
ψ_1	0.071	1.926	-	-	-	
ψ_2	0.084	1.672	0.146	0.153	-	
$-2logL$	551.937		551.939		552.850	

Table 8: Estimates of μ_2 and values of $-2logL$ under different predetermined values of the non- random dropout parameter (ψ_1) as a sensitivity analysis (*: the value of the parameter estimated by the NRD model).

ψ_1	μ_2	$-2logL$
-0.50	5.639	615.501
-0.35	5.681	615.513
-0.30	5.694	615.483
-0.25	5.707	6153439
-0.20	5.720	615.383
0.00	5.763	615.107
0.05	5.771	615.041
0.10	5.778	614.884
0.15	5.783	614.937
0.20	5.787	614.904
0.272*	5.789	614.880
0.30	5.790	614.888
0.35	5.788	614.909
0.40	5.784	614.954
0.45	5.779	615.024
0.50	5.771	615.121

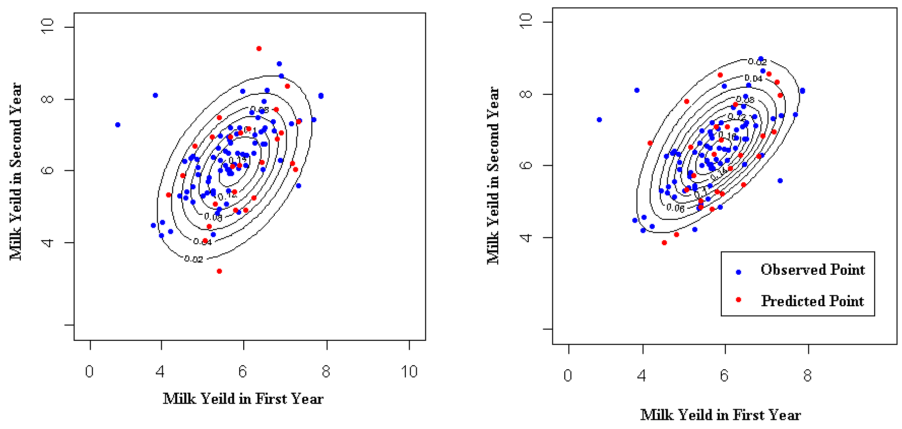

Figure 1: Mastitis data: scatter plot of the milk yield in first year versus that of the second year after imputation of missing values using generation of $f(y_2|y_1)$ along with the contour plot of the fitted bivariate skew-normal distribution for all data (right panel) and the same plot after fitting bivariate normal distribution (left panel).

4.2 Application 2: Rats Data

We will now consider a relatively small example with a three-variate response. These data come from a randomized experiment designed to study the effect of the inhibition of testosterone production in rats (Department of Orthodontics of the Catholic University of Leuven in Belgium (Verbeke and Lesaffre 1997; Verbeke and Molenberghs 2000). A total of 50 male Wistar rats were randomized to either the control or one of the two treatment groups (low or high dose of the drug Decapeptyl (triptorelin), a testosterone production inhibitor) and analyzed before this in Molenberghs and Verbeke (2001) and Verbeke and Molenberghs (2004). Treatment started at the age of 45 days, and measurements were taken every 10 days, with the first observation taken at the age of 50 days. The response measurement is a characterization of the size of the skull, taken under anaesthesia. The investigators' impression is that dropout is independent of the measurements. We consider three beginning time points and the individual profiles are shown in Figure 3. To linearize, we use the logarithmic transformation $t_{ij} = ln(1 + (Age - 45)/10)$ and we consider the following model,

$$Y_{ij} = \beta_0 + \beta_1 L_i + \beta_2 H_i + \beta_3 t_{ij} + \beta_4 L_i t_{ij} + \beta_5 H_i t_{ij} + \varepsilon_{ij} \quad (10)$$

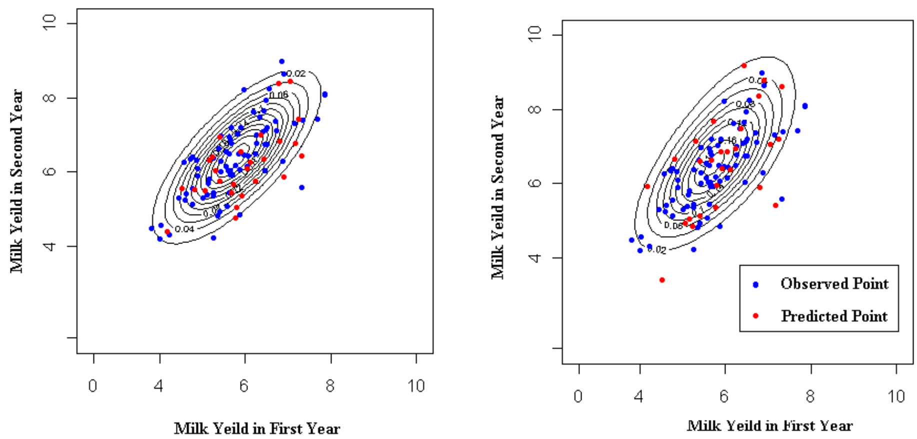

Figure 2: Mastitis data after removing outliers: scatter plot of the milk yield in first year versus that of the second year after imputation of missing values using generation of $f(y_2|y_1)$ along with the contour plot of the fitted bivariate skew-normal distribution (right panel) and the same plot after fitting bivariate normal distribution (left panel).

$$= \begin{cases} \beta_0 + \beta_1 + (\beta_3 + \beta_4)\ t_{ij} + \varepsilon_{ij} & if\ low\ dose \\ \beta_0 + \beta_2 + (\beta_3 + \beta_5)\ t_{ij} + \varepsilon_{ij} & if\ high\ dose \\ \beta_0 + \beta_3\ t_{ij} + \varepsilon_{ij} & if\ control, \end{cases}$$

where Y_i is a vector of order three containing the response values for the i^{th} animal, L_i and H_i are indicator variables such that

$$L_i = \begin{cases} 1 & if\ low\ dose \\ 0 & otherwise, \end{cases}$$

$$H_i = \begin{cases} 1 & if\ high\ dose \\ 0 & otherwise. \end{cases}$$

The joint model under skew normal distribution and normal distribution are used to obtain the parameter estimates of this model in the presence of dropout. Also, the estimates of standard errors have been obtained using the observed information matrix . The results are shown in Table 9. Results show that the use of multivariate skew normal gives a NRD mechanism. On the other hand the multivariate skew normal gives a CRD mechanism, but strong significance of skewness parameters λ_1 and

λ_3. Also according to Table 9, in skew-normal model the interaction of time and low dose of the drug, furthermore, the interaction of time and the high dose of the drug are not significant, while these two parameters are significant in normal one. In this manner, while the high dose of the drug is significant in skew normal model, is not significant under normal assumption.

Table 9: Parameter estimates and standard errors of the fit of the multivariate skew normal distribution on the rats data.

	Skew Normal Scenario		Normal Scenario	
Parameter	Est.	S.E.	Est.	S.E.
β_0	70.525	1.849	67.294	1.226
β_1	0.150	0.879	0.191	0.738
β_2	-2.554	1.909	-0.538	0.702
β_3	4.841	1.783	6.809	3.482
β_4	0.091	0.633	3.273	1.033
β_5	-0.035	1.546	2.360	0.913
σ_{11}	9.541	2.902	19.865	5.949
σ_{12}	3.987	2.112	7.002	2.828
σ_{13}	1.499	1.709	5.607	4.580
σ_{22}	5.934	2.168	8.140	2.995
σ_{23}	4.307	1.649	3.122	1.913
σ_{33}	6.315	3.087	8.752	2.709
λ_1	-3.080	1.711	-	-
λ_2	4.091	2.329	-	-
λ_3	5.409	1.723	-	-
ψ_0	0.042	1.484	-0.598	0.697
ψ_1	-2.295	2.526	-3.769	1.102
ψ_2	2.214	1.518	3.602	1.193
$-2logL$	636.380		674.098	

5 Conclusion

In this paper, we concentrated on using of a new family of distribution named as skew-normal in dropout problems. At first we used this distribution in some sets of simulated data and studied the effects of different factors, like rate of missingness and sample size, in estimating the parameters in a bivariate skew-normal distribution. Our study was under a NRD framework applying the non-ignorable model, which is the most

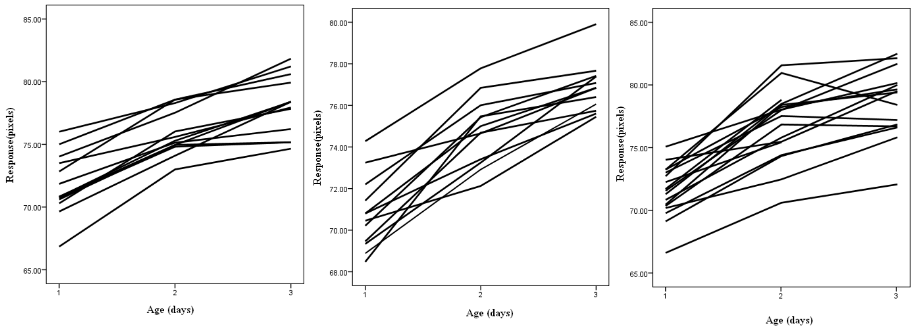

Figure 3: Rats data: Individual growth curves for each of the treatment groups in the rats experiment data.

popular approach in analysing incomplete data studies. Our simulation studies showed that increasing sample size and decreasing missing rate decrease bias of parameter estimates and standard deviations.

We also used a bivariate skew-normal to study and analyze well-known Mastitis data. Diggle and Kenward (1994) had studied this problem. Crouchley and Ganjali (2002) also studied it in two situations: one with outliers and another without outliers, they used a selection model and multivariate normal distribution in their study and obtained a NRD and RD mechanism in analyzing whole data and data without outliers, respectively. We used bivariate skew-normal distribution for the two responses and a probit regression as a model for missingness mechanism to analyze the data. Considering all data we found a RD mechanism and considering data without outliers we found a CRD mechanism. We found that the skew parameters is significant and not considering this skewness may mislead the analysis to have a NRD mechanism. In analyzing another three-variate response data we reached to the same conclusion.

References

Arellano-Valle, R. B., Del Pino, G., and San Matrin, E. (2002), Definition and probabilistic properties of skew-distributions. Statist. Probab. Lett., **58**, 111-121.

Arellano-Valle, R. B., and Genton, M. G. (2005), On fundamental skew distributions. J. Multivariate Anal., **96**, 93-116.

Arellano-Valle, R. B., Bolfaine, H., and Lachos, V. H. (2005a), Skew-normal linear mixed models. J. Data Sci, **3**, 415-438.

Arellano-Valle, R. B., Ozan, S., Bolfarine, H., and Lachos, V. H. (2005b), Skew-normal measurement error models. J. Multivariate Anal., **96**, 265-281.

Azzalini, A. (1985), A class of distribution which includes the normal ones. Scand. J. Statist., **12**, 171-178.

Azzalini, A. (1986), Further results on a class of distributions which includes the normal ones. Statistica XLVI, 199 - 208.

Azzalini, A. (2005), The skew-normal distribution and related multivariate families. Scandinavian Journal of Statistics, **32**, 159-188.

Azzalini, A. and Capitanio, A. (1999), Statistical applications of the multivariate skew-normal distributions. J. Roy. Statist. Soc., B, 579-602.

Azzalini, A. and Dalla-Valle, A. (1996), The multivariate skew-normal distribution. Biometrika, **83**, 715-726.

Baghfalaki, T. and Ganjali, M. (2011), An EM estimation approach for analyzing bivariate skew-normal data with non-monotone missing values, Commun. Statist. Theory and Methods, **40**(9), 1671-1686.

Bayes, C. L. and Branco, M. D. (2007), Bayesian inference for the skewness parameter of the scalar skew-normal distribution. Brazilian Journal of Probability and Statistics 21: 141–163.

Bolfarine, H., Olfarine, H., and Lachos, V. (2006), Skew binary regression with measurement errors. Statistics, **40**, 485-494.

Bolfarine, H. and Lachos, V. H. (2007), Skew probit measurement error models. Statistical Methodology, **4**, 1-12.

Cancho, V. G., Lachos, V. H., and Ortega, E. M. (2008), A nonlinear regression model with skewnormal errors. Statistical Papers, 10.1007/s00362-008-0139-y.

Crouchley, R. and Ganjali, M. (2002), The common structure of several recent statistical models for dropout in repeated continuous responses. Stat. Modelling, **2**, 39-62.

Demirtas, H. (2007), Practical advice on how to impute continuous data when the ultimate interest centers on dichotomized outcomes through pre-specified thresholds. Communications in Statistics - Simulaion and Computation 36:871-889.

Diggle, P. J. and Kenward, M. G. (1994), Informative dropout in longitudinal data analysis (with discussion). Appl. Stat., **43**, 49-94.

Diggle, P. J., Heagerty, P., Liang, K. Y., Zeger, S. L. (2002), Analysis of Longitudinal Data. Oxford: Oxford University Press.

Ganjali. M and Ranji, H. (2008), A Comparison of Several Algorithms and Models for Analyzing Multivariate Normal Data with Missing Responses. Appl. Math., **3**, 55-68.

Greenlees, W. S., Reece, J. S., and Zieschang, K. D. (1982), Imputation of missing values when the probability of response depends on the variable being imputed. Journal of the American Statistical Association, **77**, 251-261.

Glynn, R. J., Laird, N. M., and Rubin, D. B. (1986), Selection modeling versus mixture modeling with nonignorable nonresponse. in H. Wainer (ed.), Drawing inferences from self selected samples, 115-142, New York: Springer-Verlag.

Hardin, J. W. and Hilbe, J. M. (2003), Generalized Estimating Equations. Florida: Chapman and Hall.

Heckman, J. J. (1979), Sample Selection Bias as a Specification Error. Econometrica. **47**, 153-161.

Henze, N. A. (1986), A probabilistic representation of the skew-normal distribution. Scand. J. Stat., **13**, 271-275.

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

Kenward, M. G. (1998), Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. Statist. Med., **17**, 2723-2732.

Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B., and Montenegro, L. C. (2007), Likelihood based inference for multivariate skew-normal regression models. Commun. Statist. Theory and Methods, **36**, 1769-1786.

Leigh. J, Ward. M, and Fries. J, (1993), Reducing attrition bias with an instrumental variable in a regression model: results from a panel of rheumatoid arthritis patients. Statistics in Medicine, 12: 1005-1018.

Liang, K. Y. and Zeger, S.L. (1986), Longitudinal data analysis using generalized linear models. Biometrika, **73**, 1322.

Lin, I., Ho, J., and Chen, L. (2009), Analysis of multivariate skew-normal models with incomplete data. JMVA, **100**, 2337-2351.

Liseo, B. and Loperfido, N. (2004), Default Bayesian analysis of the skew-normal distribution. J. Statist. Plann. Inf, **136**, 373 - 389.

Little, R. J. A. (1993), Pattern-mixture models for multivariate incomplete data. J. Amer. Stat. Associat., **88**, 125-134.

Little, R. J. A. (1994), A class of pattern-mixture models for normal incomplete data. Biometrika, **81**, 471-483.

Little, R. J. A. (1995), Modelling the dropout mechanism in repeated measures studies. JASA, **90**, 1112-1121.

Little, R. J. A. and Rubin, D. B. (2002), Statistical Analysis with Missing Data. New York: John Wiley & Sons, Inc.

Molenberghs G. and Kenward M. (2007), Missing Data in Clinical Studies. New York : John Wiley & Sons, Inc.

Molenberghs, G., Thijs, H., Kenward, M. G., and Verbeke, G. (2003), Sensitivity analysis of continuous incomplete longitudinal outcomes. Statistica Nederlandica, **57**, 112-135.

Molenberghs, G. and Verbeke, G. (2001), A review on linear mixed models for longitudinal data, possibly subject to dropout. Statistical Modelling, **1**, 235-269.

Molenberghs G., Verbeke, G., Thijs, H., Lesaffre, E., and Kenward, M. G. (2001), Mastitis in Dairy Cattle: Influence Analysis to Assess Sensitivity of the Dropout Process. Comput. Stat., **37** 93-113.

Rubin, D. B. (1976), Inference and missing data. Biometrika, **63**, 581-592.

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.

Sartori, N. (2006), Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. Journal of Statistical Planning and Inference, 136:4259–4275.

Verbeke, G. and Lesaffre, E. (1997), The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Computational Statistics and Data Analysis, **23**, 541556.

Verbeke, G. and Molenberghs, G. (2000), Linear Mixed Models for Longitudinal Data. New York: Springer.

Verbeke, G. and Molenberghs, G. (2004), Longitudinal and incomplete data Analysis. Proceedings of the international biometric conference, Cairns, Australia.

Wu, M. C. and Bailey, K. R. (1988), Analysing changes in the presence of information right censoring caused by death and withdrawal. Statistics in Medicine, **7**, 337-346.

Wu, M. C. and Bailey, K. R. (1989), Estimation and comparison of changes in the presence of information right censoring: conditional linear model. Biometrika, **45**, 939-955.

Wu, M. C. and Carroll, R. J. (1988), Estimation and comparison of changes in the presence of informative censoring by modeling the censoring process. Biometrics, **44**, 175 - 188.

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