Mathematical Statistics
Stockholm University

Marginal regression model using inverse probability weights for handling dropouts missing at random in longitudinal data

Yen Ngo

Examensarbete 2006:3

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Examensarbete **2006:3**,
http://www.math.su.se/matstat

Marginal regression model using inverse probability weights for handling dropouts missing at random in longitudinal data

Yen Ngo*

April 2006

Abstract

Missing data occur invariably in longitudinal studies. Subjects may drop out before the study terminates, or be lost to follow-up in such a way that no further measurements are provided after the time of dropout. Statistical methods which ignore the mechanism for dropout will lead to biased inference. Generally, which method is to be considered for handling incomplete data, depends on which type of dropout mechanism it is. As the focus in this report is on dropouts missing at random, i.e., the probability of dropout is related to the observed responses, inverse probability weights (IPW) approach is being applied to account for dropouts. The underlying idea behind IPW methodology is, each available observation at a particular occasion is given a weight that is inverse cumulative probability of being observed at that time. To describe how inverse probability weights can be applied in marginal regression model, we use the longitudinal data on body weights from 550 rats who were randomized in a clinical trial to receive a daily dose of new substances of concentration 0 or 1.0 , 5.0 and 25-36 mg/kg. The primary interest lies on comparison of control and dose groups. However, the analysis of body weight data is complicated by dropouts due to death, in the sense that rats with low body weight at one measurement occasion tend to drop out of the trial at the next occasion. Based on this longitudinal data, results from IPW approach will be compared with those obtained from an 'unweighted' analysis.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: fk03yng@math.su.se. Supervisor: Joanna Tyrcha.

Preface

This report is a master thesis in mathematical statistics and is done at Stockholm University and AstraZeneca in Södertälje.
I would like to thank my supervisor, Niclas Sjögren at AstraZeneca, not only for his contribution to an interesting thesis but also for giving useful comments concerning statistical methods using in this report and for his great help in SAS programming.
I would also like to thank my supervisor, Joanna Tyrcha, at Department of Mathematical Statistics, Stockholm University, for her support and help in Latex.

Finally, I want to highlight that handling dropouts in longitudinal clinical studies is a difficult issue and to everyone who is unfamiliar to this broad area, it can be very confused to address, for example, the distinction between different types of dropout mechanisms and the choice among existing applicable methods. Last, but not least, a special word of thanks to professor Garrett M. Fitzmaurice, at Department of Biostatistics, Harvard University, Boston, for his advice that helped me in right direction.

Contents

1 Introduction **4**

2 General linear models for longitudinal data **5**

2.1 Notation 5

2.2 Covariance matrix 7

2.3 Multivariate normal distribution 7

2.4 Parameter estimation 8

2.5 Residual maximum likelihood estimation 10

2.6 Newton - Raphson Iteration 11

2.7 Different types of tests 12

2.7.1 Wald test on single parameters 12

2.7.2 Wald test on groups of parameters 13

2.7.3 Likelihood ratio test 14

3 Marginal models **14**

3.1 The link function 16

3.2 Covariance structures for repeated measures 16

3.2.1 Unstructed covariance 16

3.2.2 Compound symmetry 17

3.2.3 Autoregressive: AR(1) 17

3.2.4 Toeplitz 18

3.2.5 Heterogeneous AR(1) 18

3.3 Selection of model for covariance 19

3.3.1 Comparing nested models for the covariance 19

3.3.2 Comparing non-nested models for the covariance 20
3.4 Sandwich estimator 20
3.5 Marginal model for a continuous response 22
3.6 Estimating marginal models 22

4 Dropouts **23**
4.1 Missing at random (MAR) 24
4.2 Inverse probability weights 25

5 Application **27**
5.1 Clinical trial of some substances in rats 27
5.2 Purposes of the report 27
5.3 Model formulation 28
5.3.1 Weight caculation for MAR analysis 30
5.3.2 Marginal regression model via inverse weighting 33
5.3.3 Discussion 42

1 Introduction

In a longitudinal clinical study, each individual or subject is measured repeatedly on the same outcome at a number of time points. A common problem with analyzing longitudinal data is not all responses are observed at all occasions, resulting in a large class of patterns of missing values. In general, there are two distinguishable types of missing data: intermittent, in which an individual's response can be missing at one follow-up time and then be measured at the later follow-up time, and dropouts, in which once an individual drops out from the study prematurely, no subsequent repeated measures of the response are obtained on that individual. The latter is referred as monotone missing data pattern. Individuals drop out from the study for various reasons, include death, adverse reactions, lack of improvement, recovery, undesirable side effects and other factors unrelated to specific treatment or outcome.

In the presence of dropouts, the standard methods of longitudinal analysis may yield biased estimates and standard errors. Thereby the choice of analytical method has important implications on the estimates of the outcome variable and the relationship between outcome and covariates. Thus the mechanisms producing the missing observations must be considered. Following the terminology described by Little and Rubin [11] and Fitzmaurice et al [5] for longitudinal data, dropout mechanisms can be classified into one of three categories. First, an individual is dropping out completely at random (MCAR) if the probability of dropout at each occasion is independent of both the observed and the missing values of the response. Second, an individual is dropping out at random (MAR) if the probability of dropout at each occasion is related to the observed responses, but not the missing values. A more detail of MAR is outlined in section 4.1. Third, a dropout mechanism is non-ignorable if the probability of dropout at each occasion depends on unobserved responses.

There is a large literature on statistical methodology for handling different types of dropout mechanisms mentioned above. When inferences about the average response in the subpopulation sharing a common covariate vector are the focus, fitting marginal models is the most appropriate. One of the chief attractions with marginal models is it does not require complete specification of the joint distribution of the longitudinal response but rather is based only on specification of the means and variances of the responses. When the data are MCAR, the marginal approach will give consistent estimates for the parameters. While under MAR or non-ignorable assumptions, the estimating equations are not unbiased and they fail to provide consistent estimates. Inverse probability weighted estimating equations described in Robins et al [14] have been developed to deal with biases that may result from incomplete data which are MAR. Here the underlying idea is, each observed response is given a weight that is inverse cumulative probability of being observed in order to adjust for dropouts.

The main focus of this report is on the application of marginal regression model using inverse probability weights to adjust for dropouts from a clinical trial of some substances of a new drug that are tested in rats. In the trial, 234 of 550 rats dropped out due to death, leaving behind an amount of missing measurements. The missing responses (body weights) are assumed to be MAR, since rats who are observed to be weaker (via their previous observed body weights) are more likely to drop out when they reach a certain value of the body weight, as long as their probability of dropout does not further depend on their missing responses.

In the foregoing, the report is structured as follows. A review of general linear models, including distributional assumption, parameter estimation and statistical inferences is outlined section 2. Marginal model for continuous response with different covariance patterns is considered briefly in section 3. Dropouts and weighting approach are set out in section 4. Section 5 describes the application of the weighting method in which, the results are compared to those in an unweighted method. Conclusions and discussion are also involved in this section.

2 General linear models for longitudinal data

The term ' longitudinal data ' as used in this report refers to the data in form of repeated measurements on the same unit (human, plan, plot, sample etc.) over time. That is, the same response is measured again and again on the same individual under a fixed time interval. The scientific questions of interest in this kind of data, often involve not only the usual questions, such as how the mean response differs across treatments, but also how the change in the mean response over time differs and other issues concerning the relationship between response and time. Thus, it is necessary to present the situation in term of a statistical model that acknowledges the way in which the data were collected in order to address these questions. In the following, we will study the ways to model these data and explore the approaches to analyzing them. To begin, we introduce a general linear model with response variable that is continuous and assumed to have an approximate multivariate normal distribution.

2.1 Notation

Following the description of linear models of Fitzmaurice et al [5] for longitudinal data, we assume that there are n individuals and each individual i is to be measured repeatedly on a set of t_i times ($i = 1, \ldots, n$). The notation t_i allows different individuals to have unequal numbers of observations, which implies the possible presence of missing values. We can form t_i x 1 response

vector

$$Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{it_i} \end{pmatrix}$$

where Y_{ij} is the outcome variable for the ith individual at time j, i =1,. . . ,n and j =1,. . . ,t_i. Each observation Y_{ij} has an associated set of covariates that may be either regression-type continuous variables or dummy variables indicating class membership. Let X_{ij} denote the vector of covariates with length p.

$$X_{ij} = \begin{pmatrix} X_{ij1} \\ X_{ij2} \\ \vdots \\ X_{ijp} \end{pmatrix}$$

Next we consider a linear regression model for Y_{ij} , which is modeled as a linear function of covariates as

$$Y_{ij} = \beta_1 X_{ij1} + \beta_2 X_{ij2} + \ldots + \beta_p X_{ijp} + e_{ij} \tag{1}$$

where $\beta = (\beta_1, \beta_2, \ldots \beta_p)$ is a $p \times 1$ vector of unknown regression parameters and e_{ij} are unknown independent and identically distributed normal random errors, with mean zero and variance σ^2, representing deviations of the responses from their corresponding predicted means

$$E(Y_{ij} \mid X_{ij}) = \beta_1 X_{ij1} + \beta_2 X_{ij2} + \ldots + \beta_p X_{ijp}$$

With longitudinal data, we expect e_{ij} to be correlated within individuals. That is

$$Cov(e_{ij}, e_{ik}) \neq 0 \qquad (j \neq k)$$

The preceding equations in (1) can be written using vectors and a matrix as follows:

$$\begin{pmatrix} Y_{i1} \\ Y_{i1} \\ \vdots \\ Y_{it_i} \end{pmatrix} = \begin{pmatrix} X_{i11} & X_{i12} & \ldots & X_{i1p} \\ X_{i21} & X_{i22} & \ldots & X_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{it_i1} & X_{it_i2} & \ldots & X_{it_ip} \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ \ldots \\ e_{it_i} \end{pmatrix}$$

Note that the first column of the matrix X_i consists of ones corresponding to the intercept β_1, for all i and j. For simplicity and extendibility, this entire system can be presented in form

$$Y_i = X_i\beta + e_i,$$

where $Y_i = (Y_{i1}, \ldots, Y_{i2}, \ldots, Y_{it_i})'$ denote the $t_i \times 1$ vector of responses, X_i is $t_i \times p$ matrix of covariates and $e_i = (e_{i1}, e_{i2}, \ldots, e_{it})'$ is an $t_i \times 1$ vector of random errors.

2.2 Covariance matrix

We have noted previously that in longitudinal study, each individual is observed at several points over time. Measures on different individuals are independent, while repeated measures on the same individual are not, i.e. correlated, so covariance concern is only with the latter. The covariance structure refers to the variances at individual times and to correlation between measures at different times. There are basically two aspects of the correlation. The first one is, two measures on the same individual are correlated simply because they share common contributions from the individual. This is due to variation between individuals. Second, measures on the same individual that are taken at adjacent times are typically more highly correlated than measurements taken several time points apart. This is covariation within individuals. Also, variances of repeated measurements are not usually constant over study time. In context to longitudinal data, heterogeneity of variance across occasions can be accounted by allowing the diagonal elements of the covariance matrix to differ.

Let Y_i be a response vector that contains t_i repeated measurements for individual i with elements $Y_i = (Y_{i1}, Y_{i2} \dots Y_{it_i})$. Then the variances and the covariances or correlation for Y_i can be combined into a covariance matrix

$$Cov \begin{pmatrix} Y_{i1} \\ Y_{i1} \\ \vdots \\ Y_{it_i} \end{pmatrix} = \begin{pmatrix} Var(Y_{i1}) & Cov(Y_{i1}, Y_{i2}) & \dots & Cov(Y_{i1}, Y_{it_i}) \\ Cov(Y_{i2}, Y_{i1}) & Var(Y_{i2}) & \dots & Cov(Y_{i2}, Y_{it_i}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_{it_i}, Y_{i1}) & Cov(Y_{it_i}, Y_{i2}) & \dots & Var(Y_{it_i}) \end{pmatrix}$$

Of note, we allow $Cov(Y_i)$ to be "unrestricted" or " unstructed", which means it can take on any valid pattern of variances and correlations (as long as it is a proper covariance matrix, i.e. that it is positive definite). An overview of different covariance structures is described in section 3.2.

2.3 Multivariate normal distribution

In general linear models, the vector of continuous responses Y_i is often assumed to have a multivariate normal distribution with mean

$$E(Y_i) = \mu_i = X_i\beta$$

and covariance matrix

$$Cov(Y_i) = \Sigma_i$$

The multivariate normal probability density function has the following representation

$$f(y_i) = (y_{i1}, y_{i2}, \dots, y_{it_i}) = (2\pi)^{t_i/2} \mid \Sigma_i \mid^{1/2} exp\{-\frac{1}{2}(y_i - X_i\beta)'{\Sigma_i}^{-1}(y_i - X_i\beta)\}$$

where $\mid \Sigma_i \mid^{1/2}$ is the *determinant* of Σ_i (also known as the *generalized variance*).
Some well known properties of multivariate normal distribution are

- $f(y_i)$ is completely determined by $\mu_i = X_i\beta$ and Σ_i
- and depends in a large extent on $(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta)$

The latter has interpretation in terms of a measure of *distance*.

2.4 Parameter estimation

Estimation of the unknown parameters β and Σ_i is often done using the standard method of *maximum likelihood* (ML). The idea behind ML estimation is to determine values of β and Σ_i that best predict the observed data. These values are obtained by maximizing a so-called *likelihood function*. The likelihood function is defined by using the density function of the observations. Normally, with independent observations, the likelihood function is simply the product of the individual univariate normal density functions for Y_{ij}. In the case where observations on the same individual are not independent, i.e. correlated, the likelihood function has to be based on the joint density function for the vector of repeated measures. To find the maximum likelihood estimate of β in the setting of correlated data we first make the assumption that Σ_i is *known*. Given that $Y_i = (Y_{i1}, Y_{i2} \ldots Y_{it_i})'$ is multivariate normal distributed, the likelihood function can be written in the form

$$L = \prod_{i=1}^{n} (2\pi)^{t_i/2} \mid \Sigma_i \mid^{1/2} exp\{-\frac{1}{2}(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta)\}.$$

To derive the ML-estimate β we have to maximize L with respect to β. For the normal distribution, an extreme value will always be a maximum. The maximum value of likelihood function will coincide with the maximum value of the logarithm of the same likelihood function and if we take the logarithm of the likelihood function, we will replace the product sign with a summation sign which makes derivation somewhat easier. Thus, instead of maximizing L we could equally well maximizing the logarithm of the likelihood function. The latter is denoted l and is given by

$$l = -\frac{K}{2} log(2\pi) - \frac{1}{2}\sum_{i=1}^{n} log \mid \Sigma_i \mid -\frac{1}{2}\sum_{i=1}^{n}(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta)$$

where $K = (\sum_{i=1}^{n} t_i)$ is the total number of observations. Since we are maximizing l with respect to β, all terms that do not include β are eliminated. In addition, the third term of log-likelihood function has a negative sign which

means that we minimize

$$\sum_{i=1}^{n}(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta) \tag{2}$$

Differentiation of (2) with respect to β yields

$$\widehat{\beta} = \{\sum_{i=1}^{n}(X_i'\Sigma_i^{-1}X_i\}^{-1}\sum_{i=1}^{n}(X_i'\Sigma_i^{-1}Y_i)$$

The resulting value $\widehat{\beta}$ is called the *generalized least square estimate* (GLS). Suppose that Σ_i is known, we can then outline some properties of GLS estimator of β. For any choice of Σ_i, the GLS estimate of β is unbiased

$$E(\widehat{\beta}) = (\sum_{i=1}^{n}X_i'\Sigma_i^{-1}X_i)^{-1}\sum_{i=1}^{n}X_i'\Sigma_i X_i\beta = \beta$$

and has covariance

$$Cov(\widehat{\beta}) = \{\sum_{i=1}^{n}X_i'\Sigma_i^{-1}X_i\}^{-1}$$

In large samples with the meaning that the sample size n, grows larger while number of repeated measures and model parameter remains fixed, the distribution of β has the form

$$\widehat{\beta} \sim N(\beta, \{\sum_{i=1}^{n}(X_i'\Sigma_i^{-1}X_i\}^{-1})$$

So far we have based estimation of β on the assumption that Σ_i is known but usually Σ_i is unknown and has to be replaced by an estimate. Estimation of Σ_i is done in similar manner as with estimation of β. That is we obtain estimate of Σ_i by solving the derivative of the log-likelihood function with respect to Σ_i and then finding the values of Σ_i that make those derivatives equal to 0. However, this equation is non-linear and has to be solved by using numerical algorithms that maximize the likelihood. A commonly used algorithms is Newton-Raphson (see section 2.6). Once the iterative estimate of Σ_i, say $\widehat{\Sigma}_i$, has been found, it is used to caculate the GLS-estimate of β,

$$\widehat{\beta} = \{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}Y_i) \tag{3}$$

In larges samples, the resulting estimator of β in (3) will have all the same properties as when Σ_i is known.

2.5 Residual maximum likelihood estimation

Restricted or residual maximum likelihood estimation (REML) is an alternative method to find the best unbiased estimator of Σ_i since it is a well known fact that ML-estimates of Σ_i are downward biased in small samples (e.g. the diagonal elements of Σ_i are underestimated). To get an idea of what REML is and how it differs from ML, we will consider the case where observations are independent, i.e. each observation Y_{ij} is univariate normal distributed with mean, $\mu_{ij} = X'_{ij}\beta$ and constant variance, σ^2. Thus, the log likelihood function is defined as

$$l = -\frac{K}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{t_i}(Y_{ij} - X_i\beta)^2$$

Maximizing l with respect to σ^2 gives

$$\widehat{\sigma}^2 = \sum_{i=1}^{n}\sum_{j=1}^{t_i}(Y_{ij} - X'_{ij}\widehat{\beta})^2/K$$

where $K = n \times t_i$. The expected value of σ^2 is

$$E(\widehat{\sigma}^2) = (\frac{K-p}{K})\sigma^2$$

where p is the dimension of β. This indicates that ML-estimate of σ^2 is biased if the number of observations is small relative to the number of parameters and underestimates σ^2. The bias arises because the ML-estimate has not regarded the fact that β, also, is estimated. That is, in the estimator of σ^2 we have replaced β by $\widehat{\beta}$. This has led to the development of REML, where unbiased estimator of σ^2 can be obtained. The basic idea behind REML is to eliminate the parameters β from the likelihood so that it is defined only in terms of σ^2. This can be done by using the likelihood for the residuals which will depend only on σ^2, and not on β. Suppose that a linear regression for Y_{ij} is

$$Y_{ij} = X_{ij}\beta + e_{ij}$$

where e_{ij} are residuals (errors) and univariate normal distributed with mean zero and constant variance σ^2. Then the ML-estimate of β is

$$\widehat{\beta} = \{\sum_{i=1}^{n}\sum_{j=1}^{t_i} X'_{ij}X_{ij}\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{t_i}(X_{ij}Y_{ij})$$

We can estimate the residuals by

$$\widehat{e}_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{t_i}(Y_{ij} - X_{ij}\widehat{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{t_i}(Y_{ij} - X_{ij}(X'_{ij}X_{ij})^{-1}X_{ij}Y_{ij})$$

The ML-estimate of σ^2 is computing the variance of the estimated residuals:

$$\widehat{\sigma}^2 = \frac{1}{K}\sum_{i=1}^{n}\sum_{j=1}^{t_i}\widehat{e}_{ij}^2$$

In terms of algebra, the ML-estimate of σ^2 will be

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{\widehat{e}'_{ij}\widehat{e}_{ij}}{K} = \frac{1}{K}(Y_{ij} - X_{ij}\widehat{\beta})'(Y_{ij} - X_{ij}\widehat{\beta}) \\ &= \frac{1}{K}(Y_{ij} - X_{ij}(X'_{ij}X_{ij})^{-1}X_{ij}Y_{ij})'(Y_{ij} - X_{ij}(X'_{ij}X_{ij})^{-1}X_{ij}Y_{ij})\end{aligned}$$

This estimator of σ^2 is too small or downward biased because

$$E(\widehat{\sigma}^2) = E(\frac{\widehat{e}'_{ij}\widehat{e}_{ij}}{K}) = \frac{K-p}{K}\sigma^2$$

where p is dimension of X_{ij}. The unbiased estimator of the variance of the estimated residuals, which is the REML estimator, is

$$\begin{aligned}s^2 &= \frac{\widehat{e}'_{ij}\widehat{e}_{ij}}{K-p} \\ &= \frac{(Y_{ij} - X_{ij}(X'_{ij}X_{ij})^{-1}X_{ij}Y_{ij})'(Y_{ij} - X_{ij}(X'_{ij}X_{ij})^{-1}X_{ij}Y_{ij})}{K-p}\end{aligned}$$

and has the expected value $E(s^2) = \sigma^2$. Here, the REML estimator is based on $K - p$ error contrasts, that is, it is based on what is left over after we get rid of the regression coefficients β. In the setting of correlated data, where Y_i has a multivariate normal distribution with mean $\mu_i = X_i\beta$ and $Cov(Y_i) = \Sigma_i$, REML-estimate of Σ_i is obtained by maximizing the residual log likelihood

$$\frac{1}{2}\sum_{i=1}^{n} log \mid \Sigma_i \mid -\frac{1}{2}\sum_{i=1}^{n}(y_i - X_i\widehat{\beta})\Sigma_i^{-1}(y_i - X_i\widehat{\beta}) - \frac{1}{2}log \mid \sum_{i=1}^{n} X_i'\Sigma_i^{-1}X_i \mid$$

Note that the residual log likelihood includes an extra determinant term that makes a correction in a manner similar to changing the divisor as in linear regression above.

2.6 Newton - Raphson Iteration

Residual log likelihood and ML-method described in previous section have no closed-form solution, therefore an iterative algorithm like Newton-Raphson, is required to calculate estimates of Σ_i. Let $f(\theta)$ be the residual log likelihood

function above, where θ are the parameters in Σ_i that need to be estimated. From Newton-Raphson procedure a root of

$$f'(\theta) = \frac{\partial f(\theta)}{\partial \theta} = \mathbf{0}$$

can be found which is a maximum. Using only the first order Taylor series approximation to the function $f'(\theta)$ about θ_0, we have

$$f'(\theta) = f'(\theta_0) + \frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'}(\theta - \theta_0) = \mathbf{0}$$

Setting $f'(\theta)$ to $\mathbf{0}$ we obtain the root

$$\theta_1 = \theta_0 - \left[\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'}\right]^{-1} f'(\theta_0)$$

Now, θ_1 can be replaced by θ_0 and the Newton-Raphson algorithm with $(m+1)$ iteration will be

$$\theta^{(m+1)} = \theta^{(m)} - \left[\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'}\right]^{-1}\Bigg|_{\theta=\theta^{(m)}} f'(\theta^{(m)})$$

2.7 Different types of tests

Hypotheses and confidence intervals on single parameters or groups of parameters can be tested in different ways in general linear models. To test hypotheses about β we can make direct use of the ML-estimate $\widehat{\beta}$ and its estimate covariance matrix

$$\widehat{Cov}(\widehat{\beta}) = \{\sum_{i=1}^{n}(X_i'\Sigma_i^{-1}X_i\}^{-1}$$

2.7.1 Wald test on single parameters

A test of the hypothesis that single parameter β_k of vector β is zero can be made by comparing the Wald statistic

$$Z = \frac{\widehat{\beta}_k}{\sqrt{\widehat{Var}(\widehat{\beta}_k)}} \tag{4}$$

with the appropriate percentage point of the normal distribution. $\widehat{Var}(\widehat{\beta}_k)$

is the diagonal component of $\widehat{Cov}(\widehat{\beta})$ corresponding to β_k. Similarly

$$\beta = \widehat{\beta}_k \pm z_\alpha \sqrt{\widehat{Var}\widehat{\beta}_k} \tag{5}$$

would provide a $(1-\alpha)$ confidence interval for the parameter β_k.

2.7.2 Wald test on groups of parameters

Generally, it is often of interest to make inference about linear combinations of the components of β. Let L denote a vector or matrix of known weights (often representing contrasts of interest, for example treatment effects or differences between treatment groups). Here L is assumed to be a matrix whose rows represent different linear combinations for a single linear combination, that is, L is a row vector. Now, let us suppose that it is of interest to test $H_0 : L\beta = 0$ and as an example let $\beta = (\beta_1, \beta_2, \beta_3)$ and L = (0,0,1). The test $H_0 : L\beta = 0$ will be equivalent to $H_0 : \beta_3 = 0$. A natural estimate of $L\beta$ is $L\widehat{\beta}$ and the covariance matrix of $L\widehat{\beta}$ is given by $LCov(\widehat{\beta})L'$. Thus, the sampling distribution of $L\widehat{\beta}$ is

$$L\widehat{\beta} \sim N(L\beta, LCov(\widehat{\beta})L')$$

As mentioned above, L is a single row vector then $LCov(\widehat{\beta})L'$ is a single value (scalar) and the standard error of the estimator $L\widehat{\beta}$ is obtained as $\sqrt{LCov(\widehat{\beta})L'}$. An approximate confidence interval for $L\beta$ is

$$L\widehat{\beta} = L\widehat{\beta} \pm z_\alpha \sqrt{LCov(\widehat{\beta})L'} \tag{6}$$

We can also use a t critical value from a t-distribution in place of the normal critical value, with degrees of freedom chosen in various ways. When testing $H_0 : L\beta = 0$ versus $H_1 : L\beta \neq 0$, we can form the Wald test statistic

$$Z = \frac{L\widehat{\beta}}{\sqrt{LCov(\widehat{\beta})L'}} \tag{7}$$

and compare Z to the critical values of the standard normal distribution. Based on the fact that if Z is standard normal random variable, then Z^2 follows a χ^2 distribution on 1 degree of freedom. Thus, we could conduct the identical test by comparing Z^2 to the appropriate χ^2 critical value with one degree of freedom. In fact, we can write Z^2 equivalently as

$$W = (L\widehat{\beta})'(LCov(\widehat{\beta})L')^{-1}(L\widehat{\beta}) \tag{8}$$

The generalization of L having more than one row and testing more than one hypothesis is straight forward. If we want to test $H_0 : L\beta = 0$ corresponding

to $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ then H_0 can be expressed as

$$L\beta = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \beta_1 - \beta_2 \\ \beta_1 - \beta_3 \end{pmatrix}$$

If L has r rows, then a simultaneous test of the r contrasts has the same form in (6)

$$W = (L\widehat{\beta})'(LCov(\widehat{\beta})L')^{-1}(L\widehat{\beta})$$

which follows a χ^2 distribution with r degree of freedom under the null hypothesis.

2.7.3 Likelihood ratio test

Another way to test the regression parameters is the *likelihood ratio test*. The likelihood ratio test of $H_0 : L\beta = 0$ versus $H_1 : L\beta \neq 0$ is based on the following principle. Suppose that we have two models that have the same fixed effects, except one model has $L\beta = 0$. The 'full' model is the one with all the parameters and the reduced model is the one with $L\beta = 0$. That is, the reduced model is simpler than the full model, so that when the reduced model holds the full model must necessarily hold. The reduced model is said to be *nested* within the full model. We can compare two nested models by comparing their maximized log-likelihoods that are denoted by $\widehat{l}_{full}$ and $\widehat{l}_{red}$. The likelihood ratio statistic will be of the form

$$G^2 = -2log(\widehat{L}_{red}/\widehat{L}_{full}) = -2[log(\widehat{L}_{red}) - log\widehat{(L}_{full})] = -2(\widehat{l}_{red} - \widehat{l}_{full}) \quad (9)$$

where $\widehat{L}_{red}$ and $\widehat{L}_{full}$ are likelihood functions for respective model. If H_0 is true (as well as other assumptions) then the test statistic (9) is asymptotically distributed as a χ^2 random variable with degrees of freedom equal to the number of parameters in full model minus the number of parameters in reduced model. The likelihood ratio test in (9) indicates that the larger the differences between $\widehat{l}_{full}$ and $\widehat{l}_{red}$ the stronger the evidence that the reduced model is inadequate or H_0 is rejected in favor of H_1 at level of significance α if

$$G^2 > \chi^2_{r,1-\alpha}$$

where r is the difference in number of parameters in the two models.

3 Marginal models

Marginal models can be viewed as an extension of general linear models in the case of correlated data. Additionally, marginal models is the most

appropriate approach if the main interest is in population average (e.g. mean effect of treatment within a population) but also includes a model accounting for within-subject correlation among the responses. As note in Fitzmaurice et al [5], "the term marginal is used here to emphasize that the mean response modeled is conditionally only on the covariates of interest and not on any random effects or previous responses". Specially, marginal models do not require complete specification of the joint distribution of the longitudinal response, only a regression model for the mean response. That is, marginal models provide a unified approach to model all types of response variables including continuous, discrete, survival and count responses.
Let Y_i= (Y_{i1}, ..., Y_{i2}, ...,$Y_i t_i$)' denote a t_i x 1 vector of correlated responses, where Y_{ij} is the outcome variable for the i^{th} individual at time j, $i = 1,\ldots,n$ and $j = 1, \ldots, t_i$ and $X_{ij} = (X_{ij1}, X_{ij2} \ldots X_{ijp})$ is $p \times 1$ vector of covariates. With marginal models, following assumptions are made:

1. The marginal expectation or mean of the response, $E(Y_{ij} \mid X_{ij}) = \mu_{ij}$, is modeled as a function of explanatory variables, X_{ij}, through a known link function
$$g(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta$$

2. The marginal variance of each Y_{ij} is related to the marginal mean, μ_{ij} through
$$Var(Y_{ij}) = v(\mu_{ij})\phi$$
where $v(\mu_{ij})$ [1] is a known 'variance function' and ϕ is a scale or dispersion parameter that may be known or estimated.

3. The correlation between Y_{ij} and Y_{ik} is a function of the means μ_{ij}, μ_{ik} and perhaps of additional parameters, ρ, that may also need to be estimated.

More details of these assumptions will be coming shortly. It is worth to keep in mind that with marginal models the mean and the within-subject dependence are modeled separately. This separation of the modeling of the mean response and the correlation among responses has important implications for interpretation of the regression parameters in the model for the mean response. That is, the regression coefficients β, are interpreted for the population rather than for individual or more specifically, they have interpretations in term of how the effects of certain covariates on the marginal expectations or average responses vary among sub-populations (e.g., different treatment or exposure groups). Of note, the occurrence of the correlation does not alter the interpretation of β.

[1] $v(\mu_{ij})$ describes how the variance of the response is functionally related to the mean of the response.

3.1 The link function

The link function $g(.)$ is a function relating the expected values, μ_i of the response Y_i to the predictors $X_{i1}, X_{i2} \ldots X_{ip}$. It has the general form

$$g(\mu_i) = \eta_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} = X'_{ip}\beta$$

where η_i is often called 'linear predictor' which is a linear combination of the vector of unknown regression parameters $\beta = (\beta_1, \beta_2, \ldots \beta_p)'$ and the vector of covariates. The function $g(.)$ must be monotone and differentiable. For a monotone function we can define the inverse function $g(.)^{-1}$ by the relation $g^{-1}(g(\mu_i)) = \mu_i$. The choice of link function depends on the type of data. For continuous normal theory data, the link function is an identity link: $g(\mu_i) = \mu_i$ and the inverse is simply $\mu_i = g(\mu_i)$. For data in the form of counts from a Poisson distribution, μ_i is restricted to be positive and the link function is $g(\mu_i) = log(\mu_i)$ and its inverse has the form $\mu_i = exp(g(\mu_i))$. As an final example, for data in the form of proportions, μ_i is restricted to the interval [0,1] with a logit link function $g(\mu_i) = log(\mu_i/(1-\mu_i))$ and the inverse is $\mu_i = exp(g(\mu_i))/1 + exp(g(\mu_i))$.

3.2 Covariance structures for repeated measures

As mentioned earlier, the repeated measurements on the same individual in longitudinal data are correlated and have to be accounted in some way in order to get reasonable statistical tests for the parameters. Accounting for the correlations between measurement occasions usually increase efficiency or the precision with which we can estimate the regression parameters. In this section, we present a number of possible covariance or correlation structures. For these structure, the covariance can be characterized in terms of the variances and the correlations modeled by a vector of parameters ρ. The objective is to find a covariance structure that is most suitable for the data and at the same time is as simple as possible.

3.2.1 Unstructed covariance

In the Unstructed covariance matrix, no assumptions about the variance (σ_j^2) and covariance of the repeated measures (σ_{jk}) are made rather they are assumed to be completely general or have no patterns. This covariance structure brings the disadvantage of having a very large number of parameters as the number of measurements on each individual increase. With t_i measurement occasions, the "unstructed" covariance matrix has $\frac{t_i \times (t_i+1)}{2}$ parameters: the t_i variances at each time point and the $\frac{t_i \times (t_i-1)}{2}$ pair wise covariances (or

correlation),

$$Cov(Y_i) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1t_i} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2t_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{t_i1} & \sigma_{t_i2} & \dots & \sigma_{t_i}^2 \end{pmatrix}$$

When the sample size is small relative to the number of parameters that need to be estimated, the unstructed covariance structure can cause problem in the convergence of the iterative procedure. Therefore, it is often an advantage to impose a more limited structure. Some of such covariance patterns are introduced below.

3.2.2 Compound symmetry

The Compound symmetry has the simplest structure of all covariance patterns. It is typically characterized by the property that the variance, σ^2, and the correlation, $Corr(Y_{ij}, Y_{ik}) = \rho$, for all j and k are assumed to be constant over time with the restriction that $\rho \geq 0$. The structure has the matrix form,

$$Cov(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$

This does not agree with the concept that discussed above that correlations between repeated measurements are larger between measurements close in time than between measurements far part in time. This property of compound symmetry covariance pattern may not be satisfactory in many cases with longitudinal data, but there are other covariance structures to be investigated.

3.2.3 Autoregressive: AR(1)

The first-order autoregressive covariance pattern specifies homogeneous variances. It also specifies that correlations between observations on the same individual are not equal, but decrease toward zero with increasing length of the time interval between observations. Let the parameter σ^2 stand for the variance of an observation and $Corr(Y_{ij}, Y_{ik}) = \rho^{|k-j|}$ stand for the pairwise correlation between observations, for all j and k and $\rho \geq 0$. In AR(1), the correlation between measurements at times one and two is ρ, between measurements at times one and three is ρ^2, between measurements at times one

and four is ρ^3, and so on. That is,

$$Cov(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{t_i-1} \\ \rho & 1 & \rho & \dots & \rho^{t_i-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{t_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{t_i-1} & \rho^{t_i-2} & \rho^{t_i-3} & \dots & 1 \end{pmatrix}$$

Note that this structure i best suited to equally-spaced measurements.

3.2.4 Toeplitz

The toeplitz has covariances of the form,

$$Cov(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{t_i-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{t_i-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{t_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{t_i-1} & \rho_{t_i-2} & \rho_{t_i-3} & \dots & 1 \end{pmatrix}$$

and the specifications that the variance is constant over time and any pair of measurements that are equally separated in time have the same correlation. That is, correlation between measurements at times one and two will be same as correlation between measurements at times two and three, and correlation between measurements at times five and sex, and so on. Of note this structure is only appropriate when the measurements are made at equal (or approximately equal) time intervals. The toeplitz covariance structure has t_i parameters: 1 variance parameter, and $(t_i - 1)$ correlation parameters.

3.2.5 Heterogeneous AR(1)

Unlike compound symmetry, autoregressive and toeplitz covariances, the heterogeneous autoregressive covariance pattern allows variances and correlations of the observations to be different for each time unit and has $t_i + 1$ parameters (t_i variance parameters and one correlation parameter).

$$Cov(Y_i) = \begin{pmatrix} {\sigma_1}^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \dots & \rho^{t_i-1}\sigma_1\sigma_{t_i} \\ \rho\sigma_1\sigma_2 & {\sigma_2}^2 & \rho\sigma_2\sigma_3 & \dots & \rho^{t_i-2}\sigma_2\sigma_{t_i} \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & {\sigma_3}^2 & \dots & \rho^{t_i-3}\sigma_3\sigma_{t_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{t_i-1}\sigma_1\sigma_{t_i} & \rho^{t_i-2}\sigma_2\sigma_{t_i} & \rho^{t_i-3}\sigma_3\sigma_{t_i} & \dots & {\sigma_{t_i}}^2 \end{pmatrix}$$

Here the heterogeneous covariance structure is presented for the autoregressive pattern but it could be applied to the compound symmetry and toeplitz covariance structures as well.

3.3 Selection of model for covariance

Since model selection criteria for the mean requires the correct specification of the model for the covariance (e.g., confidence intervals and tests of hypotheses concerning elements of β depend critically upon the correct model for the covariance), the first task is to choose a covariance structure that provide a good fit to the data. If we choose a covariance model with too little structure (e.g., unstructed), there may be too many parameters to be estimated with the limited amount of data available. This would leave too little information available for estimating β and as a result it leads to weaker inferences concerning β. On other hand, if we choose a covariance model with too much structure (e.g., compound symmetry), there is more information available for estimating β. Then there is a potential risk of model misspecification that may lead to biased inferences concerning β. Thus, choosing among models is to some extend an 'art form', but a good dose of subjectivity is also involved. In general, there are two approaches for comparing models for the covariance matrix in order to decide which model to prefer,

1. Restricted ML (REML) when the models are *nested*.
2. *Information criteria* when they are not *nested*:
    - Akaike's Information Criterion (AIC)
    - Schwarz's Bayesian Information Criterion (BIC)

3.3.1 Comparing nested models for the covariance

A standard approach for comparing two nested models, is via the likelihood ratio tests as desribed in section 2.7.3. The likelihood ratio test is carried out by

$$G^2 = -2(\widehat{l}_{red} - \widehat{l}_{full})$$

which is compared to a chi-square distribution with degrees of freedom equal to the difference in number of covariance parameters for the two models.

3.3.2 Comparing non-nested models for the covariance

For non-nested comparisons, the covariance models can be compared in terms of *Information criteria*. The idea behind these criteria is they all start with the value of the likelihood function of a model and adjust it based on model complexity (i.e., number of parameters). A somewhat more specific statement is to say that the information criteria penalize the models with many parameters in such way that simpler model are being preferred. Two most widely used criteria are Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC).

Akaike's Information Criterion (AIC) is defined as

$$AIC = \widehat{l} - c$$

where $\widehat{l}$ is either the maximized ML or REML log likelihood and c is number of covariance parameters. Formulated this way, AIC can be used to compare models with the same fixed effects (i.e., the same model for the mean) but different covariance structures. With Proc Mixed in SAS-program, the model with smallest value of AIC is deemed best.

While Schwarz's Bayesian Information Criterion (BIC) has following definition

$$BIC = \widehat{l} - \frac{c}{2} ln(n^*)$$

where $\widehat{l}$ is either the maximized ML or REML log likelihood, c is number of covariance parameters and n^* is the number of effectives individuals, n in the case of ML and $n - p$ in the case of REML estimation (where p is the dimension of β). Again, models with smaller BIC are preferred but note that when n^* is relative large, BIC penalizes models with a greater number of covariance parameters more than AIC does.
Finally, some remarks regarding Information Criteria are

- Some computer programs report the AIC and BIC with the opposite sign: large values would then indicate a good model.
- Information criteria are only "rules of thumb" and not statistical tests.
- The different criteria may not always agree as to which covariance model is best.

3.4 Sandwich estimator

Of knowledge, statistical inferences about β are based on estimated standard errors, that are obtained under a assumed model for the covariance structure.

This approach is potentially problematic if the assumed covariance structure has been misspecified. By misspecification, we mean the situations in which, for example, a compound symmetry might be assumed but correlations in fact decline over time. Alternatively, a first-order autoregressive might be assumed but really the variance and correlation increase with time. A consequence of wrong specification of the covariances is that estimated standard error of the parameter estimates may be biased. This, in turn, may affect the conclusions drawn from the analysis, like hypothesis tests. Recall, the estimator of β is given by

$$\widehat{\beta} = \{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}Y_i)$$

where $\widehat{\Sigma}_i$ is the REML estimate of Σ_i and it has covariance matrix

$$Cov(\widehat{\beta}) = \{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}Var(Y_i)\widehat{\Sigma}_i^{-1}X_i)\{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}$$

If $Var(Y_i)$ is replaced by $\widehat{\Sigma}_i$, the REML estimate of Σ_i, $Cov(Y_i)$ can be estimated by

$$\{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}$$

This formula requires that the structure of the covariance matrix within individual is correctly specified. If the covariance has been incorrectly specified, then an alternative estimator for $Var(Y_i)$ is needed. A so-called "sandwich" estimator of $Cov(\widehat{\beta})$ was developed by Huber and White [8] to make the estimates of the standard errors more robust. The sandwich estimator of $Cov(\widehat{\beta})$ is obtained from replacing $Var(Y_i)$ by

$$\widehat{V}_i = (Y_i - X_i\beta)(Y_i - X_i\beta)$$

Thus, the "sandwich" estimator which is also known as "empirical" or "robust" estimator of $Cov(\widehat{\beta})$ is estimated by

$$Cov(\widehat{\beta}) = \{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}\widehat{V}_i\widehat{\Sigma}_i^{-1}X_i)\sum_{i=1}^{n}\{X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}$$

The remarkable thing about the sandwich estimator of $Cov(\widehat{\beta})$ is that it provides a consistent estimator of the covariance even when the model for the covariance matrix has been misspecified as long as the mean is correctly specified. In large samples the empirical covariance estimator yields correctly standard errors which lead to robust versions of hypothesis tests.

3.5 Marginal model for a continuous response

Since the continuous response is of interest in this report, other types of response variables will not be covered. In this section we will point out the connections between marginal models for a continuous response and the methods for longitudinal data analysis presented in section 2.
Let Y_{ij} be a continuous response and it is of interest to characterize how changes in mean response over time depend on the covariates. Then marginal model for Y_{ij} is given by

1. The mean of Y_{ij} is related to the covariates by an identity link function

$$\mu_{ij} = \eta_{ij} = X'_{ij}\beta$$

   Under the identity link, the expected value of the dependent variable is simply a linear function of the explanatory variables multiplied by their regression coefficients.

2. The variance of each Y_{ij}, given the effects of the covariates, is ϕ and does not depend on the mean response. That is,

$$Var(Y_{ij}) = v(\mu_{ij})\phi$$

   where $v(\mu_{ij}) = 1$ and ϕ presents the variance of the conditional normal distribution of the response given covariates. The assumption that the variance is constant over time may be unrealistic and to relax it, a separate scale parameter, ϕ_j, could be estimated at the j^{th} occasion if the longitudinal design is balanced on time.

3. The within-individual correlation among repeated responses is modelling by assuming, for example a first -order autoregressive covariance structure.

$$Corr(Y_{ij}, Y_{ik}) = \rho^{|k-j|}$$

   where $0 \leq \rho \leq 1$. Here ρ is independent of the means and is the pairwise correlations of among observations. Other specific choices of the covariance structures than autoregressive are also possible for modeling the within-individual correlation. The link function can be other link functions than identity as well.

3.6 Estimating marginal models

Generalized Estimation Equations (GEE) was introduced by Liang and Zeger [9] as methods of parameter estimation of marginal models when dealing with

correlated data. The underlying idea is to generalize the usual univariate likelihood equations by introducing the covariance matrix of the vector of responses, Y_i. The GEE methodology has become very popular especially for analysis of categorical and count outcomes, though they can be use for continuous as well. However, for linear models, i.e., marginal models with an identity link function, the generalized least squares (GLS) of β discussed in section 2 can be considered a special case of the GEE approach. Thus, the estimates of parameters in marginal model for continuous response with and identity link are

$$\widehat{\beta} = \{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}Y_i)$$

where $\widehat{\Sigma}_i$ is the REML estimate of Σ_i and

$$Cov(\widehat{\beta}) = \{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}\widehat{V}_i\widehat{\Sigma}_i^{-1}X_i)\{\sum_{i=1}^{n}(X_i'\widehat{\Sigma}_i^{-1}X_i\}^{-1}$$

where $\widehat{V}_i = (Y_i - X_i\beta)(Y_i - X_i\beta)$ is an estimate of $Var(Y_i)$ in order to obtain a robust estimator of $Cov(\widehat{\beta})$.

4 Dropouts

Missing data problems arise very often in longitudinal studies whenever one or more of the sequences of measurements is incomplete, in the sense that intended measurements are not taken, are lost, or are otherwise unavailable. A particular pattern of missingness that is common in longitudinal studies is dropout. Following the terminology described by Little and Rubin [11] and Fitzmaurice *et al* [5] for longitudinal data, three types of drop-out mechanisms can be distinguished based on how missing data processes depend on the responses. The first is missing completely at random (MCAR) mechanism, where the probability of drop out at each occasion is assumed to be independent of either the observed or unobserved responses (given the covariates). That is with completely random dropout, an individual leaves the study in a process which is unrelated to that individual's outcomes. The second type of dropout mechanism is missing at random (MAR), where dropout can depend on the observed responses, but is conditionally independent of the unobserved responses. In this process, it means that the probability of drop out at each occasion is conditionally independent of current and future responses, given the history of the responses prior to that occasion. The third type of dropout mechanism is a non-ignorable dropout mechanism, where the probability of dropout at each occasion is related to unobserved responses. Under MCAR and MAR assumptions, the data mechanism is often referred to as being 'ignorable'. Examples of MCAR include studies where patients

are lost to follow-up for reasons unrelated to their prognosis or where missing data are generated by staggered entry, MAR arise in situations where subjects drop out when they reach a specific value of the outcome variable, and missing non-ignorable in smoking cessation studies where the outcome variable is whether a person is smoking or not at a given time point, and researcher often assume that if a subject is missing at a particular time point, it is because they are smoking. Statistical methods which ignore the mechanism for drop out will lead to biased estimates and standard errors. In this section, the focus is on method for dealing with dropouts missing at random in marginal models.

4.1 Missing at random (MAR)

Dropout is missing at random (MAR) in a longitudinal clinical study if, for example, among individuals with the same set of covariates, those who are observed to be sicker via their previous observed values of the response, are more likely to drop out, as long as their probability of dropout does not further depend on their missing responses. To introduce a model for dropout we need a necessary additional notation. As before, let $(Y_{i1}, Y_{i2}, \ldots Y_{it_i})'$ be a vector of correlated responses. For individual i at occasion j, where j =1,. . . ,t_i a missing value indicator R_{ij} is defined such as

$$R_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$$

Worth to mention is dropout (specially, when dropout is due to death) gives rise to monotone missing data pattern in the sense that if Y_{ij} is missing, then $Y_{i(j+1)}, \ldots Y_{it_i}$ are also missing. Alternatively, when expressed in terms of the response missing indicators, dropout refers to the case where if $R_{ij} = 0$ then $R_{i(j+1)} = \ldots = R_{it_i} = 0$. Let $\lambda_{ij}(\alpha) = \text{pr}[R_{ij} = 0 | R_{i(j-1)} = 1, X_i, Y_{i(j-1)}, \alpha]$ stand for ith individual's probability of dropout at occasion j, given the history of all available response values observed up to $(j-1)$ and X_i is a vector of covariates. Generally it is assumed that all individuals are observed on the first occasion, that is $R_{i1} = 1$ which in turn means $\lambda_{i1}(\alpha) = 0$.

Ordinarily, $\lambda_{ij}(\alpha)$ is not known and must be estimated from the observed data by fitting a logistic regression model

$$logit\{\lambda_{ij}(\alpha)\} = \alpha Z_{ij}$$

or in reverse we have

$$\lambda_{ij}(\alpha) = \frac{e^{\alpha Z_{ij}}}{1 + e^{\alpha Z_{ij}}}$$

where α is $q \times 1$ vector of unknown parameters and Z_{ij} is a design matrix which may contains the observed responses prior to dropout, and treatment

assignment and any additional covariates that are thought likely to predict dropout. The log partial likelihood for ith individual takes the form

$$\sum_{i=1}^{n}\sum_{j=2}^{t_i} R_{i(j-1)} log\{\lambda_{ij}(\alpha)^{R_{ij}}[1-\lambda_{ij}(\alpha)]^{1-R_{ij}}\} \tag{10}$$

Differentiation of (10) with respect to α gives the estimating equations

$$S_i(\alpha) = \{\sum_{i=1}^{n}\sum_{j=2}^{t_i} R_{i(j-1)}[R_{ij}-\lambda_{ij}(\alpha)\} \tag{11}$$

Setting (11) equal to zero yields $\widehat{\alpha}$, thereafter we can obtain estimate of $\lambda_{ij}(\widehat{\alpha})$, which is $\widehat{\lambda}_{ij}(\widehat{\alpha})$.
To provide consistent estimates of parameters β in weighted method two assumptions [7] must be fulfilled in addition to MAR:

Assumption 1 (Non-zero probability of remaining in study)
Given the past history of observed responses and covariates, the probability that individual i is still in the study at time j is bounded away from zero or formally pr$[R_{ij}=1|R_{i(j-1)}=1, X_i, Y_{ij}]>\delta>0$.

Assumption 2
The probability of dropout model must be correctly specified i.e. $\lambda_{ij}(\alpha) =$ pr$[R_{ij}=0|R_{i(j-1)}=1, X_i, Y_{i(j-1)}]$.

Under MAR assumption and monotone missingness the probabilities of remaining in the study, $\pi_{ij}(\alpha)$ are calculated from

$$\pi_{ij}(\alpha) = pr[R_{ij}=1 \mid R_{i(j-1)}=1, X_i, Y_{i1}, \ldots, Y_{i(j-1)}, \alpha] = \prod_{j=1}^{t_i}\{1-\lambda_{ij}(\alpha)\}$$

4.2 Inverse probability weights

Finally, adjustments for dropouts can be made by using a variety of different weighting approaches, e.g., the propensity weighted methods described in Heyting et al [6] and the inverse probability weighted (IPW) estimating equations approaches described in Robins, Rotnitzky and Zhao [14]. Here, we will apply inverse probability weights in marginal model in order to correct the bias that is caused by dropouts missing at random. The essential idea behind IPW is, if observation i has a probability of being observed of π_{ij}, then this observation should be given weight, w_{ij}, in order to replace the missing measures due to dropouts in the analysis. The weight w_{ij} for ith individual at time j is assigned as inverse of the cumulative product of

fitted probabilities, $\widehat{w}_{ij}(\widehat{\alpha}) = (\widehat{\pi}_{i1}(\widehat{\alpha}) \times \widehat{\pi}_{i2}(\widehat{\alpha}) \times \ldots \times \widehat{\pi}_{ij}(\widehat{\alpha}))^{-1}$. To get an idea of what weights are, we adopt the example that is illustrated by James R. Carpenter et al [1] as follows. Assume that the following data has been seen,

Group:	A	B	C
Response:	1 1 1	2 2 2	3 3 3

then the average response is 2. However if we observed

Group:	A	B	C
Response:	1 ? ?	2 2 2	? 3 3

then the average response is 13/6, which is biased. To correct this bias, we first calculate the probabilities of being observed in each group corresponding to 1/3 in group A, 1 in group B and 2/3 in group C. Thereafter a weighted average, where each observation is weighted by 1/[Probability of being observed], can be interpreted as

$$\frac{1 \times \frac{3}{1} + (2+2+2) \times 1 + (3+3) \times \frac{3}{2}}{\frac{3}{1} + 1 + 1 + 1 + \frac{3}{2} + \frac{3}{2}} = 2$$

Conclusion to be drawn in this case is inverse probability weighting has eliminated the bias. In summary, the probability $\widehat{\pi}_{ij}$ that individual i is still in the study at time j should be counted as $\frac{1}{\widehat{\pi}_{ij}}$ individuals rather than one in subsequent analyses, corresponding to once for herself and $\frac{1}{\widehat{\pi}_{ij}} - 1$ times for others who drop out with the same past reponses and covariates. Note that an observation with low probability of being observed will receive a large weight.

Setting all assumptions that are made in preceding sections together, we will get valid parameter estimates in longitudinal studies with dropouts missing at random by solving the weighted estimating equations

$$\sum_{i=1}^{n} (y_i - X_i\beta)' \Sigma_i^{-1} W_i(\widehat{\alpha}) (y_i - X_i\beta) = 0 \tag{12}$$

where $W_i(\widehat{\alpha}) = diag\{\widehat{w}_{i1}(\widehat{\alpha}), \ldots, \widehat{w}_{it_i}(\widehat{\alpha})\}$, for $j = 2, \ldots, t_i$, and $\widehat{w}_{i1} = 1$. Note that (12) has similiar construction as (2), they only differ in matrix $W_i(\widehat{\alpha})$. The estiamators of the parameters in weighted marginal model for continuous response with an identity link will be of the forms

$$\widehat{\beta} = \{\sum_{i=1}^{n} (X_i' \widehat{\Sigma}_i^{-1} W_i(\widehat{\alpha}) X_i\}^{-1} \sum_{i=1}^{n} (X_i' \widehat{\Sigma}_i^{-1} W_i(\widehat{\alpha}) X_i)$$

and

$$Cov(\widehat{\beta}) = \{\sum_{i=1}^{n} X_i'\Sigma_i^{-1}W_i(\widehat{\alpha})X_i\}^{-1}(\sum_{i=1}^{n} X_i'\Sigma_i^{-1}W_i(\widehat{\alpha})W_i(\widehat{\alpha})'X_i)\{\sum_{i=1}^{n} X_i'\Sigma_i^{-1}W_i(\widehat{\alpha})X_i\}^{-1}$$

5 Application

5.1 Clinical trial of some substances in rats

Next we consider the application of the aforementioned methods for handling dropouts in longitudinal clinical trial. The methods are applied to data from the clinical trial of some substances of a new drug that are tested in rats under a period of two years. In the trial, the rats were randomly selected in five groups: three doses (low, medium and high) of substances and two controls. Each rat was given a daily dose and the dose in the high dose group was increased from 25 mg/kg till 36 mg/kg on day 225. At the start of the experiment each group consisted of 55 males and 55 females (see table 1). The response of interest in this report is body weight. The rats were weighted weekly during the period of week two to seventeen, afterward they were weighted every two weeks. The rats who completed the study were totally weighted 61 times. One notable feature of the trial is that rats dropped out due to death. More than one third of rats dropped out before the termination of the trial; 40.0% of rats in the two control groups, 50.0% in the low dose group, 36.4% and 41.8% in the medium and high dose group, respectively dropped out. The number of deaths is different between the sexes. Table 2 shows the mortality of males and females separately. The two control groups, 1 and 5, were pooled as control group with twice number of rats as compared to the dose groups. Death is a common event in these sorts of studies and causes a substantial amount of missing data on body weight which in turn means loss of information. As a consequence, missing body weights reduce efficiency and precision with which we can estimate the changes in the mean body weight. In fact, the greater the amount of missing body weights the greater the decrease in precision. Also, missing body weights give rise to bias and hence they lead to misleading inferences about changes in the mean response.

5.2 Purposes of the report

The primary purposes of this report are

- to consider statistical approaches for correcting the bias that caused by deaths of the rats and how these approaches affect the bias.

- to investigate the dose -response relationship, i.e. if the treatments have effect on body weight.
- to compare dose groups with control group in differences of the treatment effects based on different types of tests and confidence intervals.

5.3 Model formulation

In the proceeding, we analyze the data from the clinical trial described earlier by formulating two marginal models: an unweighted marginal model based on the data with missing body weights and compare its results to those from a weighted model, where missing body weights due to loss of rats have been corrected. Recall that in this report, estimating the average difference between dose groups is one of the main objectives. Thus, a marginal model seems appropriate for the applicable data. Because the response variable of interest at each occasion is the body weight and continuous, the identity link function and the scale parameter ϕ will be used. Let Y_{ij} denote the body weight from rat i at week j for $i = 1, 2, \ldots, 550$ and $j = 1, 2, \ldots, 61$. Furthermore, the covariate vector for rat i at week j includes

$$
\begin{aligned}
x_{ij2} &= \begin{cases} 1 & \text{if } i\text{th rat receives a low dose} \\ 0 & \text{otherwise} \end{cases} \\
x_{ij3} &= \begin{cases} 1 & \text{if } i\text{th rat receives a medium dose} \\ 0 & \text{otherwise} \end{cases} \\
x_{ij4} &= \begin{cases} 1 & \text{if } i\text{th rat receives a high dose} \\ 0 & \text{otherwise} \end{cases} \\
x_{ij5} &= \text{weeks} \\
x_{ij6} &= \text{week}^* x_{ijp}
\end{aligned}
$$

where $p = 1, 2, \ldots, 6$ and variable week is treated as a categorical variable with levels of 61 weeks. The covariate x_{ij6} denotes the dose-by-time interac-

Table 1: Summary of groups and doses

Group	No. of rats	Compound	Inhaled daily dose(mg/kg)
1	55 M + 55 F	Air	0
2	55 M + 55 F	Low dose	1.0
3	55 M + 55 F	Medium dose	5.0
4	55 M + 55 F	High dose	25/36
5	55 M + 55 F	Air	0

Table 2: Summary of the number of survivors and deaths for each sex.

Group	Females			Males		
	Total	Survivor	Death	Total	Survivor	Death
1	110	61	49	110	66	44
2	55	24	31	55	31	24
3	55	37	18	55	33	22
4	55	30	25	55	34	21

tion. Note that the control group is considered as the reference group. Table 3 display the data from the first five rats in each group.

Table 3: Layout for repeated body weights from the first five males in each group. * denotes missing value.

Dosage (mg/kg)	Rat	Sex	Body weight(gram) at week				
			1	2	3	...	105
0	1	m	28.4	28.6	28.7	...	32.1
	2	m	29.9	31.6	31.0	...	34.4
	3	m	23.5	24.4	24.4	...	*
	4	m	29.1	30.9	31.3	...	*
	5	m	30.3	30.4	30.0	...	36.5
1.0	1	m	25.4	26.5	27.8	...	32.8
	2	m	28.4	28.4	30.1	...	35.9
	3	m	26.9	27.8	28.7	...	*
	4	m	28.9	29.8	31.2	...	36.6
	5	m	31.9	32.5	33.9	...	*
5.0	1	m	29.2	31.4	31.7	...	*
	2	m	26.8	26.9	27.3	...	33.2
	3	m	25.6	26.2	26.2	...	32.4
	4	m	26.3	26.3	26.5	...	*
	5	m	30.5	31.4	31.1	...	40.3
25/36	1	m	26.6	27.5	27.8	...	28.9
	2	m	28.6	28.4	28.7	...	33.6
	3	m	26.6	27.7	27.9	...	*
	4	m	30.0	29.4	28.3	...	*
	5	m	23.6	24.9	25.3	...	*

5.3.1 Weight caculation for MAR analysis

Next, we consider the model for dropout. Given the clinical data, it is reasonable to assume that, each rat has a probability of drop-out at each week, $\lambda_{ij}(\alpha)$. And as rats with low body weights at one measurement occasion tend to drop out of the study at the next occasion, the missing body weights are assumed to be MAR. Of note, all rats have $\lambda_{i1}(\alpha) = 0$ at first week. From the logistic regression model

$$\begin{aligned}\widehat{\lambda}_{ij}(\alpha) &= pr[R_{ij} = 0|R_{i(j-1)} = 1, x_{ij2}, x_{ij3}, x_{ij4}, x_{ij5}, Y_{i(j-1)}, x_{ij5}Y_{i(j-1)}, \alpha] \\ &= \frac{exp(\alpha_{j1} + \alpha_{j2}x_{ij2} + \ldots + \alpha_{j6}Y_{i(j-1)} + \alpha_{j7}x_{ij5}Y_{i(j-1)})}{1 + exp(\alpha_{j1} + \alpha_{j2}x_{ij2} + \ldots + \alpha_{j6}Y_{i(j-1)} + \alpha_{j7}x_{ij5}Y_{i(j-1)})}\end{aligned}$$

we can estimate, for examples, $\lambda_{i2}(\alpha)$ and $\lambda_{i3}(\alpha)$ for first male who had a low dose as

$$\begin{aligned}\widehat{\lambda}_{12}(\alpha) &= pr[R_{12} = 0|R_{i1} = 1, x_{112}, x_{125}, Y_{11}, x_{125}Y_{11}, \alpha] \\ &= \frac{exp(\alpha_{21} + \alpha_{22}x_{112} + \alpha_{25}x_{125} + \alpha_{26} * 28.4 + \alpha_{27}x_{125} * 28.4)}{1 + exp(\alpha_{21} + \alpha_{22}x_{112} + \alpha_{25}x_{125} + \alpha_{26} * 28.4 + \alpha_{27}x_{125} * 28.4)}\end{aligned}$$

$$\begin{aligned}\widehat{\lambda}_{13}(\alpha) &= pr[R_{13} = 0|R_{12} = 1, x_{122}, x_{135}, Y_{12}, x_{135}Y_{12}, \alpha] \\ &= \frac{exp(\alpha_{31} + \alpha_{32}x_{122} + \alpha_{35} * x_{135} + \alpha_{36} * 28.6 + \alpha_{37} * x_{135} * 28.6)}{1 + exp(\alpha_{31} + \alpha_{32}x_{122} + \alpha_{35}x_{135} + \alpha_{36} * 28.6 + \alpha_{37}x_{135} * 28.6)}\end{aligned}$$

which imply that, given the body weights 28.4 and 28.6, the conditional probability of drop-out at second and third week are independent of the data at first and second week, respectively. Once the probabilities of missingness are fitted, we obtain the marginal probabilities that $R_{ij} = 1$ for $i = 1$ at week j= 1, 2, 3 by

$$\begin{aligned}\widehat{\pi}_{11}(\alpha) &= 1 - \widehat{\lambda}_{11}(\alpha) = 1 \\ \widehat{\pi}_{12}(\alpha) &= (1 - \widehat{\lambda}_{11}(\alpha)) \times (1 - \widehat{\lambda}_{12}(\alpha)) \\ \widehat{\pi}_{13}(\alpha) &= (1 - \widehat{\lambda}_{11}(\alpha)) \times (1 - \widehat{\lambda}_{12}(\alpha)) \times (1 - \widehat{\lambda}_{13}(\alpha))\end{aligned}$$

The weights that are inverses of the cumulative probabilities of being observed at time j are caculated in the example of three weeks as follows

$$\begin{aligned}\widehat{w}_{11}(\widehat{\alpha}) &= \widehat{\pi}_{11}(\widehat{\alpha})^{-1} \\ \widehat{w}_{12}(\widehat{\alpha}) &= (\widehat{\pi}_{11}(\widehat{\alpha}) \times \widehat{\pi}_{12}(\widehat{\alpha}))^{-1} \\ \widehat{w}_{13}(\widehat{\alpha}) &= (\widehat{\pi}_{11}(\widehat{\alpha}) \times \widehat{\pi}_{12}(\widehat{\alpha}) \times \widehat{\pi}_{32}(\widehat{\alpha}))^{-1}\end{aligned}$$

The weights for the remaining rats are calculated in similar way. Table 4 shows the inverse probability weights for the first five male rats in each

Table 4: Inverse probability weights of the first five males in each group. * denotes missing weight.

Dosage (mg/kg)	Rat	Sex	Inverse probability weight at week				
			1	2	3	...	105
0	1	m	1.00	1.00	1.00	...	1.85
	2	m	1.00	1.00	1.00	...	2.37
	3	m	1.00	1.00	1.00	...	*
	4	m	1.00	1.00	1.00	...	*
	5	m	1.00	1.00	1.00	...	2.63
1.0	1	m	1.00	1.00	1.00	...	1.85
	2	m	1.00	1.00	1.00	...	1.96
	3	m	1.00	1.00	1.00	...	*
	4	m	1.00	1.00	1.00	...	2.09
	5	m	1.00	1.00	1.00	...	*
5.0	1	m	1.00	1.00	1.00	...	*
	2	m	1.00	1.00	1.00	...	2.39
	3	m	1.00	1.00	1.00	...	2.40
	4	m	1.00	1.00	1.00	...	*
	5	m	1.00	1.00	1.00	...	3.82
25/36	1	m	1.00	1.00	1.00	...	2.10
	2	m	1.00	1.00	1.00	...	2.13
	3	m	1.00	1.00	1.00	...	*
	4	m	1.00	1.00	1.00	...	*
	5	m	1.00	1.00	1.00	...	*

group. As being seen in the table, the weights at the first three weeks and even few weeks later, for all doses are equal to one. This seems reasonable in the sense that it was start of the study and the doses might not have an effect on the body weight yet. Another ground is, we base the probability of remaining in the study at a particular week on body weight at previous week and as rats were under growth period, the cumulative probabilities of being alive are high at the first weeks which result in low weights. Remember, only available body weights at time j are given inverse probability weights in order to account for the missing measures at that time.

Before involving the fitted weights in marginal model, we have to examine the distribution of weights. Reason for this is under weighted marginal model, estimation of the parameters can become unstable when the sampling probabilities $\widehat{\pi}_{ij}(\alpha)$ are very close to zero, which leads to outsized weights.

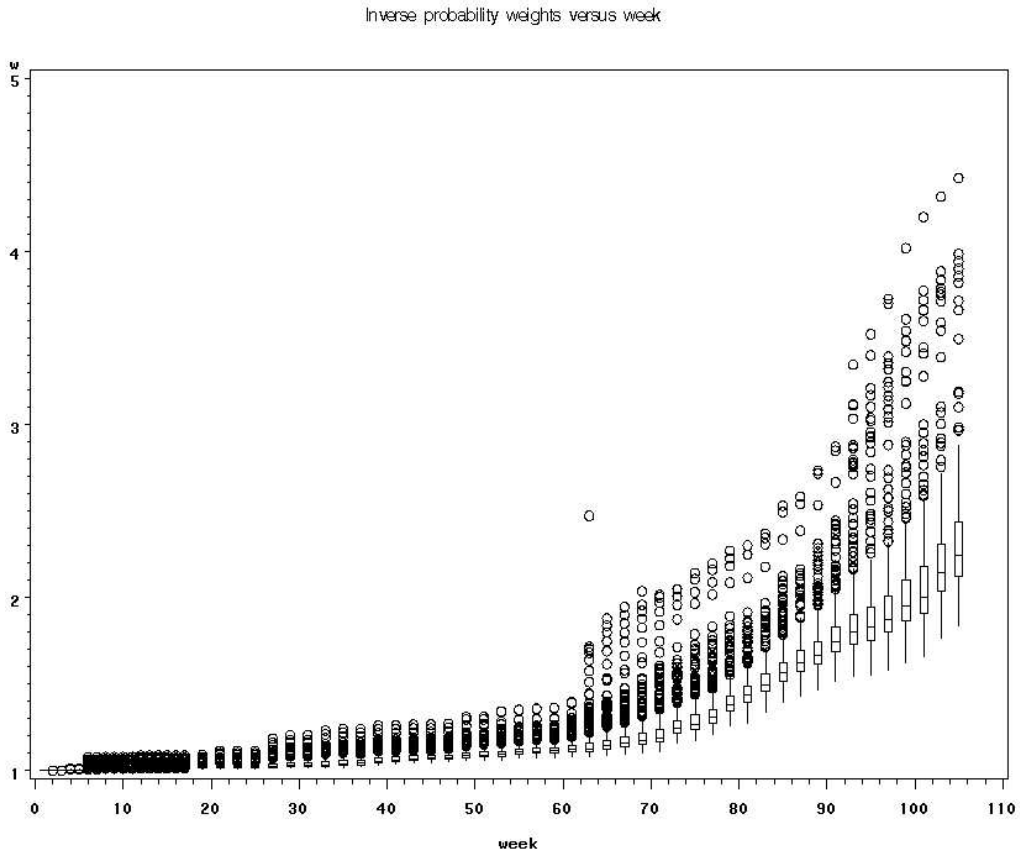

Figure 1: Box plot: Inverse probability weights vs week

When this happens, careful attention has to be paid because of these extreme weights (in the right-hand tail) may cause undue influence to available individual observations. Figure 1 displays box plot of $\widehat{w}_{ij}(\alpha)$ at each time point. The variation in weights gets larger with time, but most weights take value between 1 and 3, the maximum is around 4.5. Therefore, the estimated $\widehat{\pi}_{ij}(\alpha)$ are bounded well away from zero, with minimum value around 0.20 and large sample inference can be expected to be stable.

5.3.2 Marginal regression model via inverse weighting

Now, when the weight distribution has been checked, we can formulate the weighted model by defining

$$Y_{ij}^* = \widehat{w}_{ij}(\alpha) Y_{ij}$$

$$x_{ij}^* = \widehat{w}_{ij}(\alpha) x_{ij}$$

and the mean response model will be

$$E(Y_{ij}^* \mid x_{ij}^*) = \mu_{ij}^* = \beta_1 + \beta_2 x_{ij2}^* + \beta_3 x_{ij3}^* + \beta_4 x_{ij4}^* + \beta_5 x_{ij5}^* + \beta_6 x_{ij6}^* \qquad (13)$$

where β_1 is the population average intercept and $\beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ is the average rate of change in each x_{ijp}. In addition to this marginal model, we need to model the covariance structure of the correlated body weights on a given rat. Ideally, in the case of balance data, we would use several different covariance structures and then assess the Information Criterion or the likelihood ratio test in order to decide which structure is the best fit for the data. However, in application of IPW and owing to large number of missing body weights in unweighted analysis, only first order-autoregressive AR(1) and compound symmetry covariances can be implemented, the other structures severe computational problem. Apparently, AR(1) is our choice of covariance structure since it shares the common characteristic in longitudinal data: the correlation between measurements on the same subject to be unequal but decrease as measurements get farther apart from each other in time. In contrast, CS allows the within-subject correlation to be homogeneous over time. Moreover, in comparison of the two aforementioned approaches, the sexes are treated equally. From model (13), the parameter estimates are found as the root of the weighted estimation equations

$$\sum_{i=1}^{n} (Y_i - X_i\beta)' \Sigma_i^{-1} W_i(\widehat{\alpha}) (Y_i - X_i\beta) = 0 \qquad (14)$$

where Y_i and X_i are vectors of body weights and covariates respectively, for ith rat and $W_i(\widehat{\alpha})$ is a diagonal matrix contains inverse probability weights for ith rat. In the following tables, we summarize the results of the unweighted and weighted analyses. The interaction group*week in table 5, indicates the comparison of the dose groups in terms of their patterns of change over time

Table 5: Wald tests of fixed effects from unweighted and weighted models.

Effect	Unweigted			Weighted		
	df	Chi-square	p-value	df	Chi-square	p-value
Group	3	9.97	0.0188	3	8.53	0.0362
Week	60	2986.92	0.0001	60	2529.44	0.0001
Group*Week	180	468.51	0.0001	180	467.48	0.0001

Table 6: Unweighted marginal regression parameter estimates and standard errors (std error) for the body weight.

Effect	Estimate	Standard error	p-value
Intercept	33.0839	0.2772	0.0001
Low dose	0.1470	0.6231	0.8136
Med dose	1.2659	0.6531	0.0531
High dose	-1.8715	0.6380	0.0035
Week1	-6.9239	0.4319	0.0001
Week91	0.4757	0.2423	0.0496
Week103	-0.2026	0.1003	0.0433
Low dose*week1	0.1084	0.7292	0.8819
Medium dose*week1	-1.2405	0.7551	0.1004
High dose*weel1	1.9444	0.7421	0.0088
Low dose*week91	0.3093	0.3943	0.4327
Medium dose*week91	-0.4501	0.4315	0.2968
High dose*week91	0.4293	0.4163	0.3025
Low dose*week103	0.4058	0.1650	0.0139
Medium dose*week103	0.1382	0.1183	0.4513
High dose*week103	0.0683	0.1749	0.6939

in mean body weight. This interaction yields a Wald statistic of 468.51 and 467.48 in respective model. When compared with the reference chi-square distribution with 180 degrees of freedom, there is strong evidence to conclude that the patterns of changes over time in mean body weight are not the same in the four groups. Given the pattern of observed body weight, this result is expected. As also being seen in table 5, both group and week effects are statistically significant at 5% level as well. The weighted method gives a somewhat higher p-value in the test of the main effect of the group than the unweighted method. Based on these conclusions, we go on with taking a closer look at the parameter estimates. Since the week variable consists of

Table 7: Weighted marginal regression parameter estimates and standard errors for the body weight.

Effect	Estimate	Standard error	p-value
Intercept	33.0809	0.2772	0.0001
Low dose	0.0906	0.4943	0.8545
Medium dose	1.4109	0.4668	0.0026
High dose	-1.7113	0.4803	0.0004
Week1	-7.0278	0.3980	0.0001
Week91	0.0418	0.1866	0.0274
Week103	-1.1683	0.0727	0.0208
Low dose*week1	0.1306	0.6928	0.8505
Medium dose*week1	-1.3686	0.6840	0.0454
High dose*weel1	-1.7886	0.6884	0.0094
Low dose*week91	0.3379	0.3200	0.2911
Medium dose*week91	-0.4400	0.3215	0.1712
High dose*week91	0.4086	0.3235	0.2066
Low dose*week103	0.3770	0.1270	0.0030
Medium dose*week103	0.1169	0.1274	0.3590
High dose*week103	0.0182	0.1289	0.9269

too many levels (61 levels), we only report the estimates of three times points. In addition, it is not of primary interest to compare the dose effects at each week. The estimates from the weighted approach in table 7 imply that there is a difference between medium dose and control group, such that rats who received a medium dose of the substances have higher body weights than rats in control group. Also, rats who have been given a high dose weighted less than those from control group. Moreover, there is no difference between low dose and control group. In contrast, results from the unweighted approach (table 6) suggest that low dose respective medium dose group has the same effect on body weights as control group. While control and high dose group differ in the same sense as in the weighted analysis: rats who had a high dose weighted less than rats who only have been given the air. Furthermore, most of the parameter estimates and standard errors from the weighted analysis are smaller than those from the unweighted analysis. Table 8 shows that, rats in the high dose group have the lowest overall least square mean body weight, but the value is a little bit higher in weighted analysis. Standard errors that are obtained from weighted method are also higher, which lead to wider 95% confidence intervals. The least square means of body weight in each group, for each model are figured below. Now, we have seen that inverse probability weighting method has made some improvements, we will apply it in the following to model the sexes separately. Because of we were not able to compute other covariances than the AR(1), the AR(1) covariance

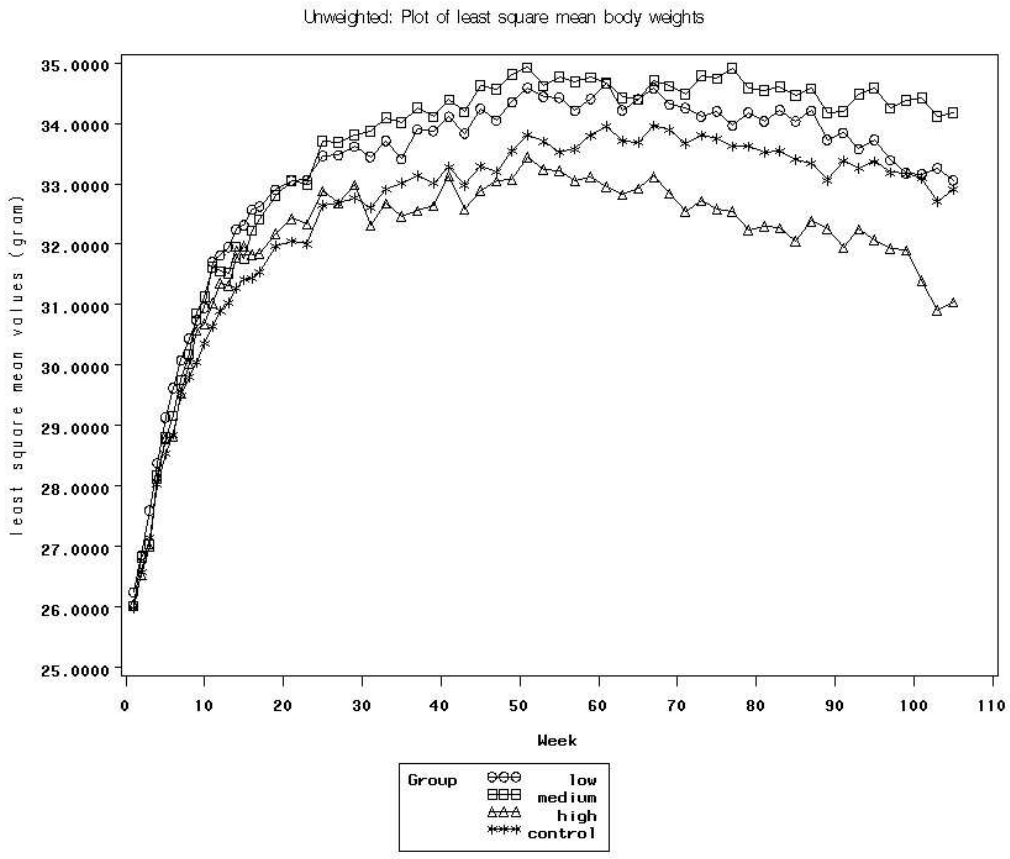

Figure 2: Unweighted: Plot of least square means of body weight

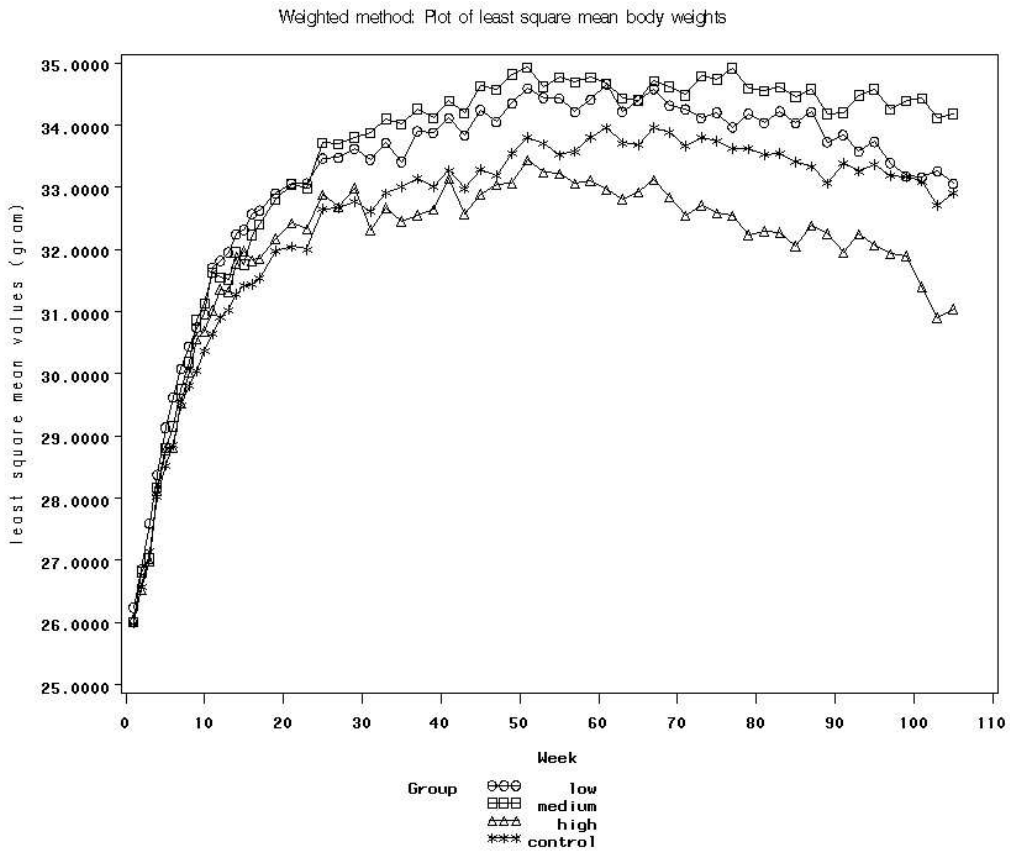

Figure 3: Weighted: Plot of least square means of body weight

Table 8: Unweighted and weighted method: Least square means of body weight and 95%-confidence interval (CI) in each dose group.

	Unweighted regression			
Effect	Estimate	Std error	95%CI	
Low dose	32.8900	0.3497	32.2031	33.5768
Medium dose	33.1345	0.3498	32.4475	33.8216
High dose	31.7741	0.3472	31.0922	32.4561
Placebo	32.2334	0.2471	31.7480	32.7187
	Weighted regression			
Low dose	32.9806	0.3881	32.2181	33.7430
Medium dose	33.3365	0.3754	32.5990	34.0739
High dose	31.9507	0.3785	31.2073	32.6942
Placebo	32.3629	0.2687	31.8350	32.8907

Table 9: Differences of least square means between control and dose groups and 95%-confidence interval (CI) for each model.

Effect	Estimate	SE	t-value	p-value	95% CI	
			Unweighted			
Low - control	0.6566	0.4282	0.53	0.1257	0.1845	1.4977
Medium - control	0.9012	0.4283	2.10	0.0358	0.05996	1.7424
High - control	-0.4592	0.4261	-1.08	-1.2963	-1.3841	0.3778
			Weighted			
Low - control	0.6177	0.4721	1.31	0.1913	-0.3096	1.5450
Medium - control	0.9736	0.0.4617	2.11	0.0354	0.06669	1.8805
High - control	-0.4121	0.4642	-0.89	0.3750	-1.32391	1.8805

may be misspecified. Hence the Sandwich estimator (SE) of $Cov(\widehat{\beta})$ is also used to provide robust standard errors. The next tables display the results of each sex. Table 10 only displays the first two rows and seven columns of the AR(1) matrix. As indicated, the autoregressive estimates show the general trend of correlations decreasing with length of time interval. In addition, the values of variance and correlations are higher for males than for females.

Conclusions drawn from tables 11, 12 and 13 are, the changes in mean body weight over time vary by group for each sex. For females, there are no differences between the control group and respective dose groups at 5% level of significance. There also seems to be no significant interaction effects

Table 10: REML variance, covariance and correlation estimates for AR(1) structure for repeated body weights.

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
			Females			
11.8810	11.2374	10.6287	10.0505	9.5060	8.9511	8.4662
11.2374	11.8810	11.2374	10.6261	10.0505	9.4637	8.9511
			Males			
15.8024	15.2920	14.7981	14.3072	13.8451	13.3978	12.9651
15.2920	15.8024	15.2929	14.7848	14.3072	13.8450	13.3978

Table 11: Wald tests of fixed effects for each sex.

Effect	Females			Males		
	df	Chi-square	p-value	df	Chi-square	p-value
Group	3	10.80	0.0128	3	10.13	0.0175
Week	60	9878.87	0.0001	60	6773.17	0.0001
Group*Week	180	2001.10	0.0001	180	4987.22	0.0001

except for low dose and high dose group at week 103. While for males, the difference between control and medium dose group is significant at 5% level, in the meaning that rats in the control group had lower body weights than rats in the medium dose group. The interaction effects are significant at 5% level for medium dose group at week one and for high dose group at week 103. Table 14 and 15 show the overall least square means body weights and the differences of overall least square means of body weight between the control and respective dose groups. The estimates suggest that males had higher body weight than females in all groups. Also rats in medium dose group had the highest body weight while rats in high dose group had the lowest body weight for both sexes. The results are displayed graphically in figures 4 and 5 below.

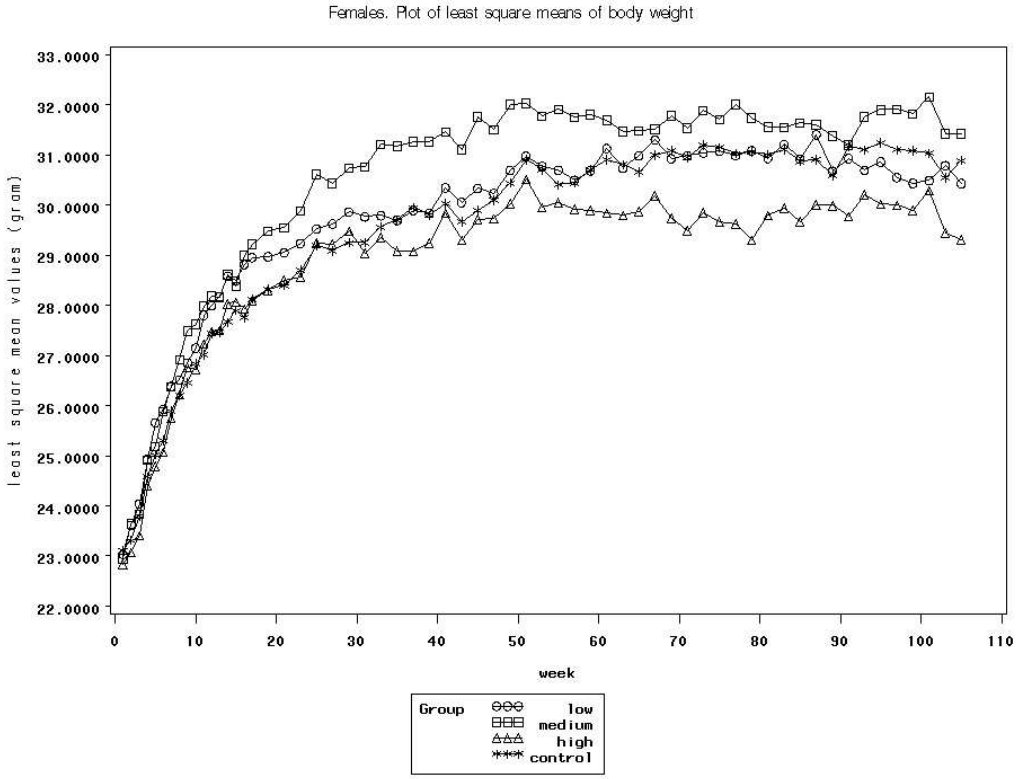

Figure 4: Females: Plot of least square means of body weight

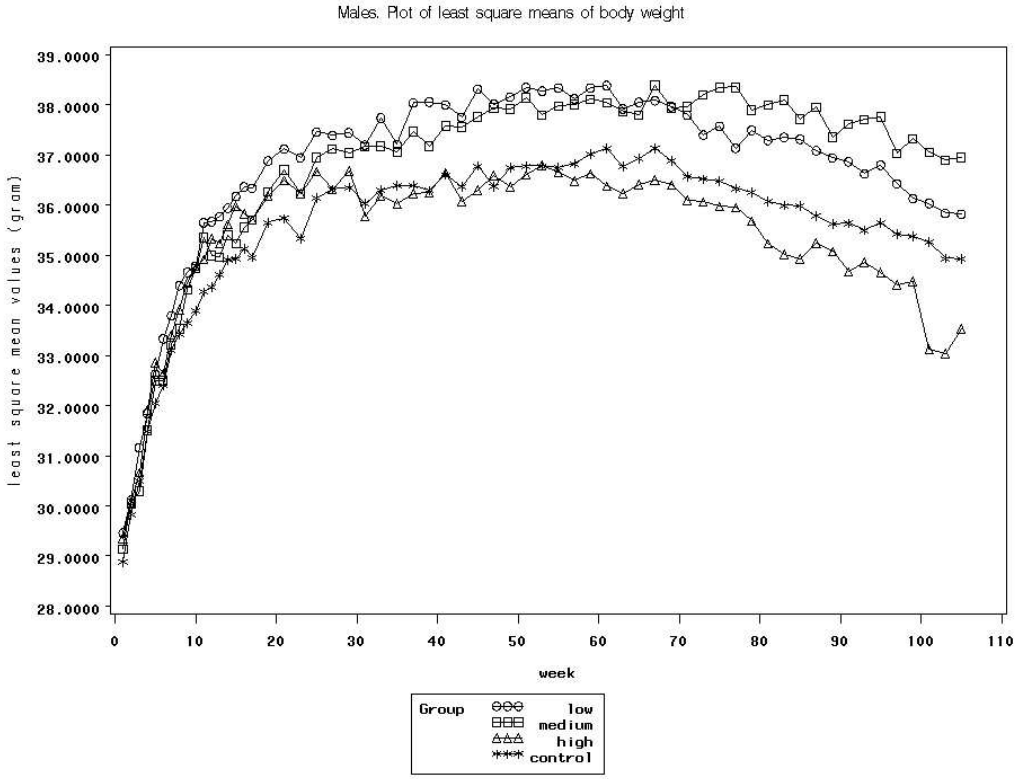

Figure 5: Males: Plot of least square means of body weight

Table 12: Weighted marginal regression parameter estimates and Sandwich estimators of standard errors for females.

Effect	Estimate	Standard error	p-value
Intercept	30.8951	0.5030	0.0001
Low dose	-0.4652	0.7661	0.5442
Med dose	0.5272	0.8889	0.5536
High dose	-1.5847	0.9095	0.0826
Week1	-7.7903	0.4492	0.0001
Week91	0.2872	0.3397	0.3979
Week103	-0.3532	0.1613	0.02860
Low dose*week1	0.3986	0.7211	0.5805
Medium dose*week1	-0.6838	0.8460	0.4190
High dose*weel1	1.3171	0.8518	0.1221
Low dose*week91	0.2125	0.4898	0.6645
Medium dose*week91	-0.5163	0.6887	0.4534
High dose*week91	0.1762	0.6005	0.7692
Low dose*week103	0.7119	0.2093	0.0007
Medium dose*week103	0.3559	0.2552	0.1632
High dose*week103	0.4869	0.2397	0.0422

5.3.3 Discussion

Inverse probability weights method is one of a few existing and common methods for handling dropouts that are missing at random and is more widely used in marginal models for discrete responses than for continuous responses. When handling with continuous outcome as in this report, the calculation of correlation between different observations on the same subject is the main difficulty. As mentioned before, we were only able to apply two covariance structures AR(1) and compound symmetry while the others caused computational problems. However, this achievement is quite satisfactory comparing to the analyses described in Dufouil et al [4], where the within-subject correlation was completely ignored in application of inverse probability weights approach. This was due to software limitation and instead a standard multiple regression using inverse probability weights was fitted, which would not be appropriate for the data in this report. The reasons are that, the correlations between body weights are strong and the numbers of measurements per rat varies markedly since some rats died very early in the trial. To ignore the correlation in our case, could give rise to inefficient parameter estimates. Another disadvantage in addition to the assumed AR(1) covariance pattern is that the scale parameter ϕ is time-invariant which is against the property of longitudinal clinical data as ours where the variance between measurements

Table 13: Weighted marginal regression parameter estimates and Sandwich estimators of standard errors for males.

Effect	Estimate	Std error	p-value
Intercept	34.9310	0.5329	0.0001
Low dose	0.8888	0.8644	0.3048
Med dose	2.02647	0.8387	0.0163
High dose	-1.4026	1.3185	0.2884
Week1	-6.0445	0.5304	0.0001
Week91	0.7062	0.3104	0.0229
Week103	0.0067	0.1396	0.9616
Low dose*week1	-0.3133	0.8853	0.7076
Medium dose*week1	-1.7717	0.7793	0.0230
High dose*weel1	1.8700	1.2353	0.1301
Low dose*week91	0.3398	0.5666	0.5487
Medium dose*week91	-0.0461	0.4768	0.9229
High dose*week91	0.4381	1.0372	0.6728
Low dose*week103	0.0167	0.2813	0.9669
Medium dose*week103	-0.0650	0.1821	0.7212
High dose*week103	-0.5012	0.1962	0.0106

Table 14: Least square means of body weight and 95%-confidence interval (CI) for each sex, in each dose group.

	Females			
Effect	Estimate	SE	95% CI	
Low dose	29.4099	0.2674	28.8834	29.9364
Medium dose	30.0980	0.3235	29.4632	30.7329
High dose	28.6423	0.3133	28.0256	29.2591
Placebo	29.1960	0.2825	28.6399	29.7521
	Males			
Low dose	36.4840	0.4282	35.6409	37.3271
Medium dose	36.4325	0.3884	35.6677	3719.72
High dose	35.1934	0.4999	34.2092	36.1776
Placebo	35.3369	0.2433	34.8579	35.8159

usually vary from occasion to occasion.

Table 15: Differences of least square means between control and dose groups and 95%-confidence interval (CI) for each sex.

Effect	Estimate	SE	t-value	p-value	95% CI	
			Females			
Low - control	0.2139	0.3890	0.55	0.5829	-0.5520	0.9797
Medium - control	0.9020	0.4287	2.10	0.0363	0.05804	1.7460
High - control	-0.5537	0.4218	-1.31	0.1904	-1.3841	0.2768
			Males			
Low - control	1.1471	0.4925	2.33	0.0206	0.1774	2.1167
Medium - control	1.0956	0.4583	2.39	0.0175	0.1932	1.9979
High - control	-0.1435	0.5560	-0.26	0.7965	-1.2381	0.9510

Finally, after accounting for missing at random dropouts, there is a slightly weaker dose-response relationship than was indicated from the unweighted analysis. It is worth to mention that in practice, it is common to report analysis only for complete-case. This approach is performed by excluding any rats that do not have measurements at all intended weeks. Hence the analysis is limited to study survivors, who tend to have higher body weight and slower decline in body weight over time than rats who die. This method is very problematic and is rarely an acceptable approach to the analysis.

References

[1] Carpenter J. R. and Kenward M. G. A comparision of Multiple Imputation and Inverse Probability Weighting for Analysis with Missing Data. *Journal of the Royal Statistical* (2005), Society, Series A.

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

[3] Diggle P. J., Liang K-J. and Zeger S. L. *Analysis of Longitudinal Data*, Oxford University Press Inc. New York 1994.

[4] Dufouil C., Brayne C. and Clayton D. Analysis of Longitudinal Studies with Death and Drop-out: a case study. *Statistics in Medicine* 2004; Vol. 23:2215-2226.

[5] Fitzmaurice G. M., Laird N. M. and Ware J. H. *Applied longitudinal analysis*, John Wiley & Sons, Inc., Publication 2004.

[6] Heyting A., Tolboom J. T. B. M. and Essers J. G. A. Statistical Handling of Drop-outs in Longitudinal Trials. *Statistics in Medicine*; Vol. 11:2043-2061(1992).

[7] Hogan J. W., Roy J. and Korkontzelou C. Tutorial in Biostatistics: Handling Drop-out in Longitudinal Studies. *Statistics in Medicine 2004*; Vol. 23:1455-1497.

[8] Huber P. The Behavior of Maximum Mikelihood Estimates under Non-Standard Conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathemathical Statistics and Probability*, vol. 1, University of California Press; Berkeley, 1967; 221-223.

[9] Liang K.-Y. and Zeger S.L. (1986). Longitudinal Analysis using Genaralized linear models. *Biometrika*; 73, 13-22.

[10] Lipsitz S., Fistmaurice G., Molenberghs G. and Zhao L. Quantile Regression Models for Longitudinal Data with drop-outs: application to CD4 cell counts of patient infected with the human immunodeficiency virus. *Applied Statistics* 19977; 46:463-476.

[11] Little R. J. A. and Rubin D. B. *Statistical Analysis with Missing Data*, 2nd ed. New York: Wiley.

[12] Olsson U. *Generalized Linear Models*; An Applied Approach. Ulf Olsson and Studentlitteratur 2002.

[13] Preisser J. S., Lohman K. K. and Rathouz P. J. Performance of Weighted Estimating Equations for Longitudinal Binary Data with Drop-outs Missing at Random. *Statistics in Medicine*; Vol 21:3035-3054.

[14] Robin J. M., Rotnizky 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*; 94:687-712.

[15] Touloumi G., Babiker A. G., Pocock S. J. and Darbyshire J. H. Impact of Missing Data due to Drop-outs on Estimatore for Rates of Change in Longitudinal Studies: a simulation study. *Statistics in Medicine*; Vol. 20:3715-3728.

[16] Yi Grace Y. and Cook Richard J. Second Order Estimating Equations for Clustered Longitudinal Data with Missing Observations. $www.stats.uwaterloo.ca/stats_navigation/techreports/02WorkingPapers\ 2002-02.pdf$.