MULTIPLE IMPUTATION COMPARED WITH SOME INFORMATIVE DROPOUT PROCEDURES IN THE ESTIMATION AND COMPARISON OF RATES OF CHANGE IN LONGITUDINAL CLINICAL TRIALS WITH DROPOUTS

Mirza W. Ali¹ and Ohidul Siddiqui²

¹Otsuka America Pharmaceutical Inc.
2440 Research Blvd.
Rockville, Maryland 20850

²Georgetown University Medical Center
Clinical Economic Research Unit
2233 Wisconsin Avenue NW, Suite 440
Washington, D.C. 20007

Key words. Multiple imputation; Informative dropout; Pattern mixture model

Abstract

Statistical analysis based on multiple imputation (MI) of missing data when analyzing data with missing observations is gaining popularity among statisticians because of availability of computing softwares; it might be tempting to use MI whenever data is missing. An important assumption behind MI is the “ignorability of missingness.” In this paper, we demonstrate the use of MI in conjunction with random effects models and several other methods that are devised to handle nonignorable missingness (informative dropouts). We then compare the results to assess sensitivity to underlying assumptions. Our focus is primarily to estimate and compare rates of change (of a primary variable). The application dataset has a high dropout rate and has features to suggest informativeness of the dropout process. The estimates obtained under
random effects modeling with multiple imputation were found to differ substantially from those obtained by methods devised to handle informative dropouts.

1. Introduction

In longitudinal clinical trials, patients are treated over a period of time and are evaluated periodically at a number of time points. Therefore, for each patient a series of measurements are made. In most trials, the treatment period and the location and number of time points for evaluation are predetermined by design. For example, a trial might be designed in which each patient would be treated for four weeks with evaluations at baseline and at the end of each subsequent week. Hence, for each patient completing the entire treatment period five observations would be available.

A primary objective in many longitudinal clinical trials is to estimate and compare the rate of change (slope) of one or more variables between groups, e.g., between a group treated with an experimental drug and a group treated with placebo. This problem of estimating and comparing rates of change becomes complicated due to missing data resulting from patient dropout.

Various approaches are available to handle missing data in longitudinal studies. Multiple imputation (MI) is an approach that is recently gaining attention among statisticians because of availability of computing softwares; it might be tempting to use MI-based statistical analysis whenever part of the data is missing. As has been pointed out by many authors (e.g., Little and Rubin (1), Little (2), and Wu and Carroll (3)), when analyzing data with missing observations it is important to get an understanding of the “missingness mechanism” and assess sensitivity of results under differing assumptions. An important assumption behind MI is that of ‘ignorable missingness.’ But, in practice, how critical is this assumption in the analysis of longitudinal clinical data? A goal of this paper is to explore this question by comparing MI-based analysis with several methods that are specifically designed to handle a particular class of ‘nonignorable missingness.’” The data set used for this comparison has features to suggest nonignorable missingness. The methods of analysis compared are the conditional linear model (Wu and Bailey (4)), pattern mixture model (Little (2)), unweighted least squares, MI-based random effects model, random effects model, MI-based ANCOVA at week 4, and ANCOVA on last observation carried forward (LOCF) data at week 4.

The paper is organized as follows. A description of the clinical trial and the data set is given in section 2, followed by a review of missingness mechanisms in section 3. Section 4 describes the various analysis methods, and in section 5 the multiple imputation process is described. Finally, results are described in section 6, followed by a discussion section.
2. Data

The application data set was generated by a clinical trial of an antipsychotic drug. Patients with schizophrenic relapse were randomized to three treatment groups, low dose, high dose, and placebo, and were to be treated for four weeks. One of the primary efficacy endpoints in the study was the total score of the positive and negative syndrome scale (PANSS) (Kay et al. (5)). This is a 30-item scale in which each item measures a particular symptom of psychosis and is rated between 1 (symptom absent) and 7 (symptom extremely severe). Patients were evaluated on this scale at baseline and at the end of each subsequent week up to four weeks. A primary objective of the study was to compare efficacy of the drug with placebo at the end of four weeks of treatment. However, an overall 41% of the patients dropped out of the study before completing the four weeks duration. An examination of the reasons for dropout revealed that patient dropout cannot be considered ‘‘random.’’ More details on the dropout phenomenon are provided in the results section.

3. Missingness Mechanisms

To understand how the results of statistical analyses of longitudinal clinical data may be affected by missing data, it is important to distinguish between various missingness mechanisms. Adopting the notation of Little and Rubin (1), let \( Y = (Y_{\text{obs}}, Y_{\text{mis}}) \) represent the data vector for a patient where \( Y_{\text{obs}} \) is the observed part and \( Y_{\text{mis}} \) is the missing part; and let \( R \) be a vector of 1’s and 0’s with 1 representing an observed component and 0 a missing component of the data vector \( Y \). The missingness mechanism is modeled by a probability model for \( R, P(R|Y, \xi) \), which depends on \( Y \) as well as some unknown parameters \( \xi \). The missing at random (MAR) assumption is that this distribution does not depend on \( Y_{\text{mis}} \), i.e., \( P(R|Y_{\text{obs}}, Y_{\text{mis}}, \xi) = P(R|Y_{\text{obs}}, \xi) \). Little and Rubin (1) also define two general classes of missingness mechanisms with respect to likelihood-based analysis. A missingness mechanism is said to be ignorable if an analysis based on the observed data likelihood alone provides valid inferences for the model parameters even when the missingness mechanism is ignored; otherwise, it is called a nonignorable missingness mechanism. Laird (6) showed that MAR is an ignorable missingness mechanism. When missingness is nonignorable, the missingness mechanism (probability) must be modeled and inferences should be based on the joint likelihood of the observed data and the missingness mechanism. In our application situation, we note that missingness is primarily due to patient dropout, i.e., all data after dropout are missing. Wu and Carroll (3) developed a likelihood-based method for estimating and comparing rates of changes for missingness due to dropouts. Their method is designed to handle a class of nonignorable mechanisms that they term ‘‘informative right censoring.’’ They assume a linear random effects
model for the data and a probit model (as a function of the random effects) for the missingness mechanism, and define a dropout process as noninformative with respect to response parameters if the joint likelihood function can be factored into two independent parts, one corresponding to the response parameters and the other corresponding to the censoring parameters. The maximum likelihood estimates of population slopes and tests of significance for differences between slopes are derived from a marginal likelihood. A computationally simpler version of Wu and Carroll’s (3) procedure was derived by Wu and Bailey (4). In their conditional linear model approach, instead of modeling the censoring mechanism, the censoring time is used as a covariate in a regression model with individual least squares regression estimates of slopes as the dependent variable. The method is described in section 4. Little (2) defines a class of nonignorable missingness mechanisms that he termed “random-effect-dependent dropout.” It is assumed that the mean response (over time) of an individual can be modeled as a function of a set of random coefficients, $\beta$, and the probability of missingness depends on $\beta$. He proposes a class of models called “random coefficient pattern mixture models” to model longitudinal data under this missingness mechanism. This method is also described in section 4.

4. Statistical Methods

4.1. Conditional Linear Model (Wu and Bailey (4))

Suppose that the patients are randomly assigned to $k$ ($\geq 2$) groups with sample sizes $n_k$, $k = 1, 2, \ldots, K$, or with total sample size $n = n_1 + n_2 + \cdots + n_k$. By design, every patient is to be evaluated on the primary variable at $T$ time points. Let $T_i$ ($\geq 2$) be the number of measurements actually made for the $i$th patient before the patient dropped out, and $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})$ be the vector of serial measurements. If the $i$th patient has complete measurements then $T_i = T$. Let $\beta_i = (\beta_{i1}, \beta_{i2})'$ be the unobservable vector representing the true initial value and the slope of the primary variable for the $i$th patient. The notation $i \in k$ will be used to denote that the $i$th patient belongs to the $k$th treatment group. The following model is assumed:

$$Y_i = X_i \beta_i + e_i$$ (1)

where $X_i$ is the $T_i \times 2$ design matrix for the $i$th patient with 1’s in the first column and the observed measurement times in the second column, $\beta_i \sim N(B_i, \Sigma)\), $e_i \sim N(0, \sigma^2 I)$, $B_i = (B_{i1}, B_{i2})'$, $\Sigma$ is the variance-covariance matrix of $\beta_i$, assumed common for all patients, and $\beta$ and $e$ are assumed independent. In this setup, $B_{i1}$ and $B_{i2}$ are the population intercept and slope for the $k$th treatment group. Our interest is to estimate and compare the slope coefficients $B_{i2}$. In particular, letting
$B_{12}$ be the slope of a placebo group, we wish to test the hypotheses $H_k: B_{12} - B_{22} = 0, k = 2, \ldots, K$.

Let $t_{ji}$ be the time point corresponding to the last observation for the $i$th patient, and let $(b_i | t_{ji}) = (X_i'X_i)^{-1}X_i'Y_i = [(b_{1i} | t_{ji}), (b_{2i} | t_{ji})]'$ be the usual least squares estimate of $\beta$, based on the $T_i$ measurements. Under the conditional linear model, it is further assumed that the estimated individual slope, given the dropout time, is a linear function of dropout time, i.e.,

$$(b_{1i} | t_{ji} = t_i) = \gamma_{00} + \gamma_{01} + \gamma_{02} + \gamma_1 t_i + \gamma_2 V_i + \epsilon_i$$

(2)

where $V_i$ is the baseline PANSS total score, $\gamma_{01}$ and $\gamma_{02}$ are the coefficients of indicator variables (taking values 1 or 0) representing the low dose and the high dose groups, $\epsilon_i$ is a random error with mean 0 and variance $\sigma^2_{Z_i}$, and $\gamma_1$ and $\gamma_2$ are unknown regression coefficients. The above model is a simpler version of a more general model given by Wu and Bailey (4), which involves polynomial function of dropout times. Wu and Bailey (4) define a test for informativeness of dropout by $H_o: \gamma_1 = 0$ against $H_a: \gamma_1 \neq 0$. Note that Eq. (2) is the usual multiple linear regression and analysis of covariance model. In the presence of dropout, however, since the individual slopes are estimated from different numbers of measurements the variance $\sigma^2_{Z_i}$ are usually different. Hence the weighted least squares method has to be used to estimate the regression coefficients. To accomplish this we need the following variance formula:

$$\text{Var}(b_i) = \text{Var}((X_i'X_i)^{-1}X_i'Y_i) = \text{Var}(\beta_i) + (X_i'X_i)^{-1}X_i'\text{Var}(\epsilon_i)$$

$$= \sum_b + (X_i'X_i)^{-1}\sigma^2_{Z_i} = \Sigma_i$$

(3)

Denoting $\text{Var}(b_i)$ by $\Sigma_i$, $\sigma^2_{Z_i}$ is the second diagonal element of $\Sigma_i$. Model (2) is estimated by weighted least squares regression. Wu and Bailey (4) define a linear minimum variance unbiased estimate (LMVUE) of $B_{12}$ by

$$\text{LMVUE}(B_{12}) = \hat{\gamma}_{00} + \hat{\gamma}_{01} + \hat{\gamma}_1 t_k + \hat{\gamma}_2 V$$

(4)

where $\hat{\gamma}_{00}, \hat{\gamma}_{01}, \hat{\gamma}_1$, and $\hat{\gamma}_2$ are estimates of the corresponding parameters, $t_k$ is the mean dropout time for patients (including completers) in the $k$th treatment group, and $V$ is the overall mean of the baseline measurements ($V_i$) for all groups combined.

Notice that model (1) is a random effects regression model; therefore the variance components $\Sigma_B$ and $\sigma^2_Z$ can be estimated, for example, by the restricted maximum likelihood (REML) method. In this application, this was accomplished by using the PROC MIXED routine of SAS (7). Estimates of $\Sigma_B$ and $\sigma^2_Z$ were used in Eq. (3) to derive an estimate of $\sigma^2_{Z_i}$ that was then employed in running the weighted least squares regression. The same estimate of $\sigma^2_{Z_i}$ was also used in estimating the variance of UWLSE($B_{12}$) in Eq. (5) below.
4.2. Unweighted Least Squares

In this method, we still assume the model given in Eq. (1), but an estimate of $B_{12}$ is now obtained as a simple average of the individual patient specific least squares regression slopes, i.e.,

$$UWLSE(B_{12}) = \sum_{i=1}^{n} \frac{b_{i2}}{n_i}$$

(5)

with variance $\text{Var}(UWLSE(B_{12})) = \sum s_{ii}^2/n_i$, where, as before, $s_{ii}^2$ is the second diagonal element of the matrix in Eq. (3). In case of informative dropout, the estimate in Eq. (5) is an unbiased estimate of $B_{12}$ (Palta and Cook (8)).

4.3. Pattern Mixture Model (Little (2))

This is an approach to handle nonignorable missingness in which the population of patients is categorized into two or more patterns of dropouts and completers; models are fitted to each pattern, and then overall estimates of the model parameters (based on the entire population of patients) are obtained as a weighted average of the pattern-specific parameter estimates. In this application we will employ a particular type of pattern mixture model, namely the random effects pattern mixture model proposed by Little (2). This type of model was also used in the analysis of longitudinal data with dropouts by Siddiqui and Ali (9). First, each patient is categorized into one of three patterns of dropout: early dropout, late dropout, and completer. Early dropouts are the ones having only the baseline and week 1 measurements; late dropouts having baseline, week 1, week 2 and week 3 measurements; and completers having baseline through week 4 measurements. Let $Y_{ij}$ represent the PANSS total score of the $i$th patient at $j$th week ($j = 0, 1, 2, 3, 4$). The following random effects model is fitted:

$$Y_{ij} = \beta_0 + \beta_1 \text{Week}_{ij} + \beta_2 \text{Dose2}_{i} + \beta_3 \text{Dose1}_{i} + \beta_4 \text{Drop1}_{i} + \beta_5 \text{Drop2}_{i} + \beta_6 (\text{Drop1}_{i} \ast \text{Week}_{ij}) + \beta_7 (\text{Dose2}_{i} \ast \text{Drop1}_{i}) + \beta_8 (\text{Dose2}_{i} \ast \text{Drop2}_{i}) + \beta_9 (\text{Dose1}_{i} \ast \text{Drop1}_{i} \ast \text{Week}_{ij}) + \beta_{10} (\text{Dose1}_{i} \ast \text{Dose2}_{i} \ast \text{Drop1}_{i} \ast \text{Week}_{ij}) + \beta_{11} (\text{Dose1}_{i} \ast \text{Dose2}_{i} \ast \text{Drop2}_{i} \ast \text{Week}_{ij}) + \beta_{12} (\text{Dose1}_{i} \ast \text{Week}_{ij}) + \beta_{13} (\text{Dose2}_{i} \ast \text{Week}_{ij}) + \beta_{14} (\text{Drop1}_{i} \ast \text{Dose1}_{i} \ast \text{Week}_{ij}) + \beta_{15} (\text{Drop2}_{i} \ast \text{Week}_{ij}) + \beta_{16} (\text{Drop2}_{i} \ast \text{Dose2}_{i} \ast \text{Week}_{ij}) + \beta_{17} (\text{Drop2}_{i} \ast \text{Dose1}_{i} \ast \text{Week}_{ij}) + \nu_{ij} + \nu_{i1} \text{Week}_{ij} + \epsilon_{ij}$$

(6)

where

Drop1 = 1 for early dropout (weeks 2, 3, 4 missing)

= 0 otherwise,
Drop2 = 1 for late dropout (weeks 3, 4 missing) = 0 otherwise,
Dose1 = 1 for low dose = 0 otherwise,
Dose2 = 1 for high dose = 0 otherwise,
\( v_0 \) = random intercept,
\( v_{1i} \) = random slope coefficient,
\((v_0, v_{1i}) \sim N(0, \Omega) \) and \( \varepsilon_{ij} \sim \text{iid } N(0, \sigma^2) \), and \((v_0, v_{1i}) \) and \( \varepsilon_{ij} \) are independent.

The model in Eq. (6) provides a mechanism for an understanding of the differences in mean responses over time among the various subgroups of patients resulting from a cross classification of the treatment groups by dropout patterns. For example, the mean response curve for placebo completers has intercept = \( \beta_0 \) and slope = \( \beta_1 \), while for placebo early dropouts the intercept and slope are \( (\beta_0 + \beta_e) \) and \( (\beta_1 + \beta_e) \), respectively. Models for other combinations of treatments and dropout patterns can be similarly derived from Eq. (6).

The population parameter \( B_k \) (either an intercept or slope) for the \( k \)th treatment arm is defined as a weighted average of the coefficients for individual dropout patterns, i.e.,

\[
B_k = \pi^{(c)} B_k^{(c)} + \pi^{(e)} B_k^{(e)} + \pi^{(l)} B_k^{(l)}
\]

where \( B_k^{(c)} \), \( B_k^{(e)} \), and \( B_k^{(l)} \) are the coefficients corresponding to completers, early dropouts, and late dropouts, respectively, for the \( k \)th treatment arm; \( \pi^{(c)} \), \( \pi^{(e)} \), and \( \pi^{(l)} \) are the population proportions of completers, early dropouts, and late dropouts. For an estimate of \( B_k \), these population proportions are estimated by the corresponding sample proportions; and estimates of \( B_k^{(c)} \), \( B_k^{(e)} \), and \( B_k^{(l)} \) are derived from the estimates of \( \beta \)'s in model (5). Hogan and Laird (10) provide the delta method approximation for the variance of \( B_k \), which we have utilized in estimating variance of \( B_k \).

4.4. Multiple Imputation (MI)-Based Analyses

Three steps are followed in MI-based analysis. First, missing data are imputed based on an imputation model. This imputation is done several times to generate a number of imputed data sets. Secondly, usual statistical analysis (e.g., estimating treatment contrasts by fitting linear models) is performed on each imputed dataset and estimates of the desired quantities and their standard errors are obtained from each imputed dataset; and lastly, these estimates and standard errors are processed by a rule given by Rubin (11) to obtain an overall estimate and standard error, which is then used to derive an overall inferential statement, e.g., a confidence
interval or \( p \)-value. For this application, 10 imputed data sets were generated. (Details of the imputation process is given in section 5.) The methods used for MI-based analyses are the random effects model (Laird and Ware (12)) and analysis of covariance on the week 4 value of PANSS total score with baseline value as a covariate.

5. Multiple Imputation

Multiple imputation (MI) has been extensively applied to handle missing data in survey sampling (Rubin (11,13)). In this section, a brief overview of MI with reference to our application is provided. A basic assumption behind MI is that of ignorable missingness. As before, let \( Y = (Y_{\text{obs}}, Y_{\text{mis}}) \) be the complete data vector with probability density \( P(Y_{\text{obs}}, Y_{\text{mis}}|\theta) \). Rubin (13) defines a “Bayesianly proper imputation” for \( Y_{\text{mis}} \) as a realization from \( P(Y_{\text{mis}}|Y_{\text{obs}}) \), where

\[
P(Y_{\text{mis}}|Y_{\text{obs}}) = \int P(Y_{\text{mis}}|Y_{\text{obs}}, \theta) P(\theta|Y_{\text{obs}}) d\theta
\]  

(8)

On the right-hand side of Eq. (8), \( P(\theta|Y_{\text{obs}}) \) is the posterior density of \( \theta \) under an assumed prior for \( \theta \). Instead of evaluating the integral in (8), the data augmentation (DA) algorithm (Tanner and Wong (14) and Schafer (15), chapters 3 and 4) is usually used for sampling from \( P(Y_{\text{mis}}|Y_{\text{obs}}) \). The DA algorithm is described below:

1. Start with an initial value \( \theta^{(0)} \) of \( \theta \).
2. At step \( (t + 1) \), draw \( Y_{\text{mis}}^{(t+1)} \) from \( P(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}) \).
3. Now draw \( \theta^{(t+1)} \) from \( P(\theta|Y_{\text{obs}}, Y_{\text{mis}}^{(t+1)}) \)

Repeat steps 2 and 3 a “large” number of times to produce the sequence \( \{\theta^{(t)}, Y_{\text{mis}}^{(t)}, t = 1, 2, 3, \ldots\} \). The following holds due to a general result of Tanner and Wong (14): The stationary distribution of \( \{\theta^{(t)}, Y_{\text{mis}}^{(t)}, t = 1, 2, 3, \ldots\} \) is \( P(\theta, Y_{\text{mis}}|Y_{\text{obs}}) \); the stationary distribution of \( \{\theta^{(t)}, t = 1, 2, 3, \ldots\} \) is \( P(\theta|Y_{\text{obs}}) \); and, finally, the stationary distribution of \( \{Y_{\text{mis}}, t = 1, 2, 3, \ldots\} \) is \( P(Y_{\text{mis}}|Y_{\text{obs}}) \). Note that the last result guarantees that, after stationarity is reached, the values of \( Y_{\text{mis}} \) that are generated via DA are, in fact, draws from the predictive distribution \( P(Y_{\text{mis}}|Y_{\text{obs}}) \). Stationarity is usually assessed by graphical methods, i.e., a time-series plot and an autocorrelation plot of the imputation model parameters.

5.1. Implementation of DA Under the Gaussian Imputation Model for the Application Data Set

In our application data set the complete data vector can be represented as \( Y = (Y_{0}, Y_{1}, Y_{2}, Y_{3}, Y_{4}) \) giving the values of PANSS total score at baseline and weeks 1 through 4. We assume that \( Y = (Y_{\text{obs}}, Y_{\text{mis}}) \sim N(Y|\theta) \). (Here \( \theta \) contains both the mean vector and the covariance matrix.) Then at the \( i \)th iteration of DA,
(\(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}\)), also follows a multivariate normal distribution. We will assume a noninformative prior for \(\theta\) so that the posterior \(P(\theta|Y_{\text{obs}}, \ Y_{\text{mis}}^{(t)})\) for \(\theta\) at the \(t\)th iteration of DA is Normal–Inverted Wishart (Schaffer (15), p. 154). These are the two probability densities needed to implement steps 2 and 3 of the DA algorithm. Starting values of \(\theta\) were obtained via the Expectation–Maximization (EM) algorithm applied to bootstrap samples from the original dataset. Schafer’s NORM software (Schafer (16)) for generating multiple imputations was used in all of the above computations. The DA algorithm was run for 1000 iterations and the 1000th value of \(Y_{\text{mis}}\) was used as imputations. The procedure described above was repeated 10 times to generate 10 imputed datasets.

5.2. Rules for Combining Results from Multiply Imputed Data Sets

Let \(Q\) be a quantity of interest to be estimated from complete data. In this application, \(Q = \) slope coefficient of a treatment group, or \(Q = \) difference between the slope of a drug treated group and that of placebo. Let \(\hat{Q}\) be an estimate of \(Q\) with a variance estimate denoted by \(U\). With \(m\) imputations, \(m\) different versions of \(\hat{Q}\) and \(U\) are obtained. Let \(\hat{Q}^{(t)}(t) = \hat{Q}(Y_{\text{obs}}, Y_{\text{mis}}^{(t)})\) and \(U^{(t)}(t) = U(Y_{\text{obs}}, Y_{\text{mis}}^{(t)})\) be the point and variance estimates using the \(t\)th set of imputed data, \(t = 1,2, \ldots, m\). Rubin (11) gave the following rule for combining them. The multiple imputation point estimate for \(Q\) is simply the average of the complete-data point estimates,

\[
\overline{Q} = \frac{1}{m} \sum_{i=1}^{m} Q^{(i)}
\]

The variance estimate of \(\overline{Q}\) is \(T = \text{Var}(\overline{Q}) = U + (1 + m^{-1})B\), where

\[
U = \frac{1}{m} \sum_{i=1}^{m} U^{(i)}
\]

is the within-imputation variance, and

\[
B = \frac{1}{m-1} \sum_{i=1}^{m} (\hat{Q}^{(i)} - \overline{Q})^2
\]

is the between-imputation variance. Inferences are based on the approximation

\[
T^{-1/2}(Q - \overline{Q}) \approx t_v
\]

where \(t_v\) stands for a \(t\) statistic with \(v\) degrees of freedom, and

\[
v = (m - 1) \left[1 + \frac{U}{(1 + m^{-1})B} \right]^2
\]
Table 1. Completion Rates by Week and Treatment Arm

<table>
<thead>
<tr>
<th></th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>14 (22)</td>
<td>13 (20)</td>
<td>7 (11)</td>
<td>30 (47)</td>
<td>64</td>
</tr>
<tr>
<td>Low</td>
<td>9 (16)</td>
<td>5 (9)</td>
<td>8 (14)</td>
<td>35 (62)</td>
<td>57</td>
</tr>
<tr>
<td>High</td>
<td>5 (8)</td>
<td>9 (15)</td>
<td>5 (9)</td>
<td>41 (68)</td>
<td>60</td>
</tr>
<tr>
<td>Total</td>
<td>28 (15.5)</td>
<td>27 (15)</td>
<td>20 (11)</td>
<td>106 (59)</td>
<td>181</td>
</tr>
</tbody>
</table>

Thus a $100(1 - \alpha)\%$ interval estimate for $Q$ is

$$Q \pm t_{v,1-\alpha/2} \sqrt{\frac{T}{n}}$$

where $t_{v,1-\alpha/2}$ is the $(1 - \alpha/2)$th quantile of the $t$ distribution with $v$ degrees of freedom.

6. Results

Table 1 provides the patient completion rates by week and treatment arm. It is seen that the trial suffered a high dropout rate. Overall, only 106/181 (58.6%) of the patients completed the scheduled duration of four weeks of study. The highest dropout occurred in the placebo group (53%), followed by low dose (38%), and high dose (32%). Early dropout (i.e., completing only up to week 1) and late dropout (i.e., completing weeks 2 and 3) rates are 15.5% and 26%, respectively.

In Table 2, the mean change from baseline in PANSS total score at each of the four weeks of treatment along with the mean scores at baseline are given. The mean change at each weekly time point is computed based on the last observation.
carried forward (LOCF) imputation method for imputing missing values due to patient dropout, except that baseline values were not carried forward to fill in missing values. For example, missing PANSS total scores at weeks 2, 3, and 4 for a patient dropping out after week 1 evaluation were filled in by the week 1 score. There were two patients (one in the placebo and another in the low-dose group) whose week 1 values were missing but one or more subsequent measurements (at weeks 2, 3, and 4) were available. For these two patients, the week 1 values were not filled in by the baseline values. This is reflected by one less number of patients ($N = 63$ for placebo and $N = 56$ for low-dose group) at week 1 for these treatment groups. A graph of these mean change scores is provided in Figure 1. It is seen that there is a slight curvature in the growth curves, in particular in the low-dose and high-dose groups. In this paper our focus is primarily in the rates of change in responses over time, and so our models include only a linear term in time. Comparing the mean change values at week 4, it is seen that the drug-treated groups showed an improvement by as much as 16 points over baseline compared to only about 1 point improvement in the placebo group.

In our subsequent analyses we have used linear models and assumed normality of PANSS total scores. PANSS total scores range from 7 to 130, with a bell-shaped curve, and has traditionally been analyzed by analysis of variance (ANOVA) or analysis of covariance (ANCOVA) (Tran et al. (17)).

A least squares regression line (of PANSS total score on weeks on study) was computed from each patient data. In Table 3, a summary of the patient-specific
Table 3. Distribution of Least Squares Slopes by Completion Time

<table>
<thead>
<tr>
<th>Completion time</th>
<th>N</th>
<th>Mean slope</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>28</td>
<td>10.79</td>
<td>21.93</td>
</tr>
<tr>
<td>Week 2</td>
<td>27</td>
<td>1.54</td>
<td>10.76</td>
</tr>
<tr>
<td>Week 3</td>
<td>20</td>
<td>-0.75</td>
<td>6.71</td>
</tr>
<tr>
<td>Week 4</td>
<td>106</td>
<td>-4.56</td>
<td>4.28</td>
</tr>
</tbody>
</table>

least squares slopes are given. Patients who dropped out within the first or the second week had average slopes of 10.79 and 1.54, respectively. The positive sign of these estimates indicates that, on average, patients dropping out within the first two weeks had a deterioration of symptoms over baseline. On the other hand, the average slope of patients who completed the study (four weeks) is -4.56, indicating an improvement over baseline. A clearly decreasing trend in mean slope over time is noted, i.e., patients having higher slopes tended to drop out early. Thus, one could argue that the probability of a patient’s dropping out at a given time point depends on his/her slope up to that time point. This fits into the definition of informative dropout of Wu and Bailey (4), and provides a basis for choosing methods of analysis based on slopes (e.g., the conditional linear model of Wu and Bailey (4), the pattern mixture model, and the unweighted least squares method).

The parameter estimates and standard errors of the conditional linear model (Eqs. (1) and (2)) are $\gamma_{00}:(18.8, 3.4), \gamma_{01}:(-1.8, 1.2), \gamma_{02}:(-2.6, 1.1), \gamma_{1}:(-3.7, 0.6),$ and $\gamma_{2}:(-0.08, 0.03).$ The $p$-value for the test of $H_0: \gamma_1 = 0$ is 0.0001, meaning that the informativeness of the dropout process is highly significant, confirming our previous conjecture that the amount of time in the study is inversely related to the patient’s slope over time. The estimate of $\gamma_1$ was also statistically significant ($p = 0.0057$), suggesting that baseline severity of the disease (as measured by PANSS total) is highly correlated with symptom status (improvement or deterioration), a negative coefficient indicating that patients with higher severity at baseline are more likely to show improvement in their symptoms than patients with milder severity at baseline. The estimates of the group slope coefficients derived by Eq. (4) are given in Table 6.

The estimates of the Pattern Mixture model parameters are given in Table 4. Estimated models for each combination of treatment group and pattern of dropout can be derived by appropriately combining the slope and intercept coefficients from Table 4. For example, the intercept and slope coefficients for early dropout in high dose are $86.48 + 0.65 + 8.52 + 1.95 = 97.6$ and $-3.46 + 18.31 - 1.83 - 11.03 = 1.99,$ respectively. These estimated models for various patterns of dropout for each treatment group is given in Table 5. It is interesting to note that slope coefficients for early dropouts in all three groups are positive indicating
that this group of patients were deteriorating in their symptoms prior to dropout. Also notice that (except for a minor deviation in the late dropout group) there is a dose response in the estimates of the slopes (i.e., decreasing slope with increasing dose) within each pattern of dropout.

Now employing Eq. (7) for combining pattern specific estimates, we may obtain overall estimates for each treatment group. The sample proportion of dropouts (and completers) are (from Table 1) 28/181, 47/181 and 106/181 for early

<table>
<thead>
<tr>
<th>Model term</th>
<th>Coefficient</th>
<th>Standard error (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>86.4790</td>
<td>2.8315</td>
</tr>
<tr>
<td>Time</td>
<td>−3.4572</td>
<td>0.8290</td>
</tr>
<tr>
<td>Dose2</td>
<td>0.6527</td>
<td>3.7237</td>
</tr>
<tr>
<td>Dose1</td>
<td>5.2555</td>
<td>3.8334</td>
</tr>
<tr>
<td>Drop1</td>
<td>8.5210</td>
<td>5.2643</td>
</tr>
<tr>
<td>Drop2</td>
<td>−2.7264</td>
<td>4.5553</td>
</tr>
<tr>
<td>Time * Drop1</td>
<td>18.3144</td>
<td>3.7889</td>
</tr>
<tr>
<td>Dose2 * Drop1</td>
<td>1.9473</td>
<td>9.4184</td>
</tr>
<tr>
<td>Dose2 * Drop2</td>
<td>3.5866</td>
<td>6.6920</td>
</tr>
<tr>
<td>Dose1 * Drop1</td>
<td>−13.8111</td>
<td>8.0639</td>
</tr>
<tr>
<td>Dose1 * Drop2</td>
<td>−3.7733</td>
<td>6.9592</td>
</tr>
<tr>
<td>Time * Dose2</td>
<td>−1.8302</td>
<td>1.0909</td>
</tr>
<tr>
<td>Time * Dose1</td>
<td>−1.1273</td>
<td>1.1244</td>
</tr>
<tr>
<td>Time * Dose2 * Drop1</td>
<td>−11.0270</td>
<td>7.2890</td>
</tr>
<tr>
<td>Time * Dose1 * Drop1</td>
<td>−4.3966</td>
<td>6.0162</td>
</tr>
<tr>
<td>Time * Drop2</td>
<td>7.1033</td>
<td>1.6823</td>
</tr>
<tr>
<td>Time * Dose2 * Drop2</td>
<td>−3.6789</td>
<td>2.5256</td>
</tr>
<tr>
<td>Time * Dose1 * Drop2</td>
<td>−4.9544</td>
<td>2.5460</td>
</tr>
</tbody>
</table>
Table 6. Estimates of Slope Coefficients

<table>
<thead>
<tr>
<th>Method</th>
<th>Low</th>
<th>High</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond.-Linear</td>
<td>-1.88 (0.866)</td>
<td>-3.32 (0.800)</td>
<td>1.31 (0.933)</td>
</tr>
<tr>
<td>Pat.-Mix.</td>
<td>-1.87 (1.023)</td>
<td>-3.29 (1.155)</td>
<td>1.22 (0.946)</td>
</tr>
<tr>
<td>UWLS</td>
<td>-1.75 (1.024)</td>
<td>-3.95 (0.894)</td>
<td>2.85 (1.084)</td>
</tr>
<tr>
<td>RE-MI</td>
<td>-3.47 (0.783)</td>
<td>-4.31 (0.714)</td>
<td>-0.87 (0.773)</td>
</tr>
<tr>
<td>RE</td>
<td>-3.90 (0.757)</td>
<td>-4.68 (0.718)</td>
<td>-0.96 (0.774)</td>
</tr>
<tr>
<td>ANCOVA-Wk4-MI</td>
<td>-3.37 (0.793)</td>
<td>-4.51 (0.728)</td>
<td>-1.08 (0.809)</td>
</tr>
<tr>
<td>ANCOVA-Wk4-LOCF</td>
<td>-2.76 (0.704)</td>
<td>-3.94 (0.686)</td>
<td>-0.34 (0.665)</td>
</tr>
</tbody>
</table>

The standard errors are given in parentheses.

dropouts, late dropouts, and completers, respectively. For example, the overall slope estimate for the high-dose group is \((28/181) \times 2.00 + (47/181) \times (-1.86) + (106/181) \times (-5.29) = -3.27\). Table 6 depicts estimates of slopes and their standard errors obtained by seven methods. Of these, the first three methods, conditional linear (Cond.-Linear), pattern mixture (Pat.-Mixture), and unweighted least squares (UWLS), are designed to handle informative dropouts. It is seen that, except for UWLS in placebo, the estimates obtained by these methods are very similar. Note also that, under the assumption of slope-dependent dropout, the conditional linear model provides a minimum variance unbiased estimate, and we observe that the variance of this estimate is the lowest among these three methods. The next two methods employ the random effects model; in the first one (RE-MI), the random effects model was used on the multiply imputed datasets, and in the second one (RE) on the observed dataset. We notice that the results of these two methods are very similar. This can partly be explained by the fact that the multiple imputations were generated under the assumption of MAR and noninformative prior so that the imputation model was primarily determined by the observed data likelihood, which is also the likelihood used by random effects model under MAR assumption. The estimates, in most cases, are substantially lower than those of the previous three methods. In particular, although all three previous methods yielded positive estimates for placebo (meaning deterioration of symptoms), these two provided negative estimates. It is noted that under informative dropout these estimates are biased. The last two methods are based on ANCOVA applied on the week 4 data only; the first one uses the multiply imputed dataset and the second one uses the LOCF data set. The slope estimates by these methods were obtained by dividing the least squares mean (LSMEAN) of change from baseline by four, and similarly the standard errors were obtained by dividing the standard errors of the least squares means by four. Overall, the results from ANCOVA at week 4 (both MI and LOCF) are similar to those of RE-MI and RE, and are also biased.

The results of treatment comparisons with placebo are given in Table 7. We
notice that the treatment effect estimates from the MI-based and LOCF-based methods are conservative compared to the methods that are designed to handle informative dropouts. The treatment effect estimates for the high dose varies from \(-6.80\) to \(-3.43\), and they are all highly statistically significant. For the low dose, the estimates range from \(-4.60\) to \(-2.30\), and while they are all statistically significant at 0.05 level, the significance levels are not as stable as those in the high dose group. Overall, in this application, all methods provided evidence in favor of effectiveness of the drug in treating the symptoms of psychosis.

### 7. Discussion

A general problem in the analysis of longitudinal clinical trials data is that of missing observations due to patient dropout. It is widely recognized that valid statistical analysis must take into account the impact of missingness on the results of analysis. To this end, it has been emphasized that the analyst must have an understanding of the missingness mechanism, and should consider methods that will minimize the impact of missingness on analysis. Two broad classes of missingness mechanisms are that of ignorable and nonignorable missingness. Analyses based on the observed data likelihood, e.g., random effects model, and multiple imputation–based analyses are valid in case of ignorable missingness. Nonignorable missingness has been characterized by two general classes of models—selection models and pattern-mixture models. An additional characterization of nonignorability was given by Wu and Bailey (4), which they term "informative ...
censoring’’ (dropout). It has been advocated that in case of analyses with missing data it is always advisable to apply more than one method to assess sensitivity of results under various assumptions of missingness.

In this paper, our goal was to estimate the rates of change (slope) in the primary variable (PANSS total score) for each treatment group, and then assess treatment effect with respect to placebo. The study had suffered a high dropout rate. We analyzed the data under assumptions of ignorability and nonignorability to assess sensitivity of results. We first investigated the dropout process by studying the distribution of individual patient slopes by dropout time (early, late, completer). We observed that time to dropout was directly related to the mean slope. This is a dropout process for which the conditional linear model of Wu and Bailey (4) and the random effects pattern mixture model of Little (2) were proposed. In our data analysis situation we consider these two methods as more appropriate for estimating rates of change. To assess sensitivity of results to assumptions about missingness mechanisms, we also applied random effects model and multiple imputation–based analysis as well as traditional methods, e.g., ANCOVA on LOCF data. Except for LOCF, these later methods are suitable for ignorable missingness. We observed that estimates of rates of change obtained from the conditional linear model and pattern mixture model are similar but differed substantially from those obtained under ignorability assumptions. We also observe that the results from random effects with multiple imputation (RE-MI) and ANCOVA at week 4 with multiple imputation (ANCOVA-Wk4-MI) are very similar to those of random effects model and did not provide any added benefit. We note, however, that the imputation was performed under the ignorability assumption and with a noninformative prior on the mean vector and covariance matrix of the normal imputation model. Although multiple imputation is a powerful tool for handling missing data, further research to improve the imputation model in case of nonignorable dropout and possible use of informative priors is clearly indicated.

Acknowledgments

We thank Dr. J. L. Schafer for providing us with the software NORM for generating multiple imputations and for his assistance in learning to use the software, and the referees whose suggestions have contributed in improving the contents of this article.

References
