Journal of Modern Applied Statistical Methods

Volume 3 | Issue 1 Article 3

5-1-2004

A Comparison Of Methods For Longitudinal Analysis With Missing Data

James Algina
University of Florida, algina@ufl.edu

H. J. Keselman
University of Manitoba, kesel@ms.umanitoba.ca

Journal of Modern Applied Statistical Methods
May, 2004, Vol. 3, No. 1, 13-26

1538 – 9472/04/$95.00

A Comparison Of Methods For Longitudinal Analysis With Missing Data

James Algina
University of
Florida

H. J. Keselman
University of Manitoba

In a longitudinal two-group randomized trials design, also referred to as randomized parallel-groups design or split-plot repeated measures design, the important hypothesis of interest is whether there are differential rates of change over time, that is, whether there is a group by time interaction. Several analytic methods have been presented in the literature for testing this important hypothesis when data are incomplete. We studied these methods for the case in which the missing data pattern is non-monotone. In agreement with earlier work on monotone missing data patterns, our results on bias, sampling variability, Type I error and power support the use of a procedure due to Overall, Ahn, Shivakumar, and Kalburgi (1999) that can easily be implemented with SAS's PROC MIXED.

Keywords: data, mixed models, split-plot design

Introduction

A randomized parallel-groups design in which participants are randomly assigned to treatments, measured on one pretreatment occasion, and on multiple post treatment occasions, is a common design for investigating treatment effects. One challenge facing researchers who use this design is how to analyze the data when there are missing observations. Little (1995), Overall, Ahn, Shivakumar, and Kalburgi (1999), Wang-Clow, Lange, Laird, and Ware (1995), and Wu and Bailey (1989) have all suggested procedures for conducting such analyses. Algina and Keselman (2003) compared a number of these methods for designs in which two treatments $(k = 1, 2)$ are compared. They concluded that while, in principle, if one has valid information about the type of missing data, the information should be taken into account in selecting a procedure, in practice it may be wise to select a method that performs well over a wide range of methods. Based on their findings, which included empirical estimates of bias, sampling variability, variations of a procedure suggested by Wu and Bailey (1989) might be considered. The principal shortcomings of these three procedures were Type I error rates above the nominal level in some conditions and, for two of the variations, a complicated method of estimation. However, Algina and Keselman acknowledged that their study should be regarded as preliminary in that they studied a limited number of conditions.

James Algina (algina@ufl.edu) is Professor of Educational Psychology at the University of Florida. His research interests are in applied statistics and psychometrics. H. J. Keselman (kesel@ms.umanitoba.ca) is Professor of Psychology, University of Manitoba. His research interests are in applied statistics. Work on this project was supported by a grant from the Social Sciences and Humanities Research Council of Canada.

One limiting factor in the Algina-Keselman (2003) study, as well as in Overall et al. (1999), Wang-Clow et al. (1995), and Wu and Bailey (1989), was a monotone pattern for the missing data. That is, once an observation was missing for a participant, no further measurements were available for that participant. Thus, a major purpose of the current investigation was to determine whether the Overall et al. procedure would continue to perform well when the missing data did not occur in a monotone pattern. In addition, the influence of a wider variety of missing data mechanisms than were included by Algina and Keselman and the influence of planned sample size on the methods were investigated. Prior to presentation of the new results, we review missing data mechanisms and the methods we investigated.

Missing Data Mechanisms

Little (1995) reviewed several mechanisms for missing data: missing completely at random (MCAR), covariate dependent (CD), and missing at random (MAR). Following Verbeke and Molenberghs (2000), when the mechanism is not MCAR, CD, or MAR, we refer to it as missing not at random (MNAR). The variables that predict which data are missing determine whether or not the data are MCAR, CD, MAR, or MNAR. In this paper we are concerned with estimation and hypothesis testing when data are missing in a design in which participants in two treatment groups are measured on one pretreatment occasion and several post treatment occasions. In such studies, there are three types of variables that describe the participants.

The first two are the potentially observable variables. These are the measurements on the variable of interest and the covariates. The latter variables include the occasion of measurement, the treatment indicator, and any other variables that are observed prior to the onset of the treatments. The third type comprises the parameters for a subject-specific within-subject model for scores on the repeated measurements. Variables in the third type are latent variables.

When the pattern of missing data at a particular time point is unrelated to the potentially observable variables and to the latent variables, the data are MCAR. If the pattern of missing data is related only to the covariates the mechanism is CD. It should be noted that some authors (see, for example Diggle & Kenward, 1994) do not distinguish between MCAR and CD missing data mechanisms. If the pattern at a particular time point is related to previous measurements on the variable of interest and the covariates in the model, but not to the actual data values that would have been observed at that time point had there been no missing data, nor to the latent variables, the data are MAR.

Methods of Analysis

In the presentation of the methods we use the following notation: Y_{ijk}, the score for the *i*th $(i = 1,\ldots,n_k)$ of n_k subjects in the *k*th $(k = 1,2)$ group on the *j*th $(j = 1,\ldots,J)$ occasion; t_j, an index for the occasion of measurement, and t_{ik}, the index value for the last measurement occasion at which the *i*th participant in the *k*th group was observed.

All of the methods, except the endpoint method studied by Overall et al. (1999), assume that if the data were complete they would conform to the following model

$$Y_{ijk} = \beta_{0ik} + \beta_{1ik} t_j + \varepsilon_{ijk} \qquad (1)$$

where

$$\varepsilon_{ijk} \sim N\left(0, \sigma^2\right)$$

and, depending on the method for analyzing the data

$$\begin{bmatrix} \beta_{0ik} \\ \beta_{1ik} \end{bmatrix} \sim N\left(\theta_k, \mathbf{D}\right)$$

or

$$\begin{bmatrix} \beta_{0ik} \\ \beta_{1ik} \end{bmatrix} \sim N\left(\theta_k, \mathbf{D}_k\right).$$

The parameters β_{0ik} and β_{1ik} are the subject-specific intercept and slope, respectively, for the

within-subject regression of the dependent variable on time of measurement.

When participants are randomly assigned to groups and it is reasonable to assume that, for each participant, the within-subjects regression is well-described by the simple linear regression model, the test of the treatment effect focuses on the average slope (i.e., the population average) in each treatment. Specifically, to test for a treatment effect one tests whether the average slopes are equal for the treatment groups.

Mixed Model for MAR Data

One method of analysis uses equation (1) as the level-1 model in a multilevel model and the following level-2 models:

$$\beta_{0ik} = \gamma_{00} + \gamma_{01} Z_{ik} + u_{0ik} \quad (2)$$

and

$$\beta_{1ik} = \gamma_{10} + \gamma_{11} Z_{ik} + u_{1ik}, \quad (3)$$

where $Z_{ik} = 1$ if the *i*th participant is in treatment 2 and 0 otherwise. The estimate of the treatment effect is $\hat{\gamma}_{11}$ and testing $H_0 : \gamma_{11} = 0$ provides a test of the treatment effect. This procedure is known to give correct results provided the data are MCAR, CD, or MAR and, in the case of the latter two mechanisms, provided that the parameters of the missing data mechanism and the parameters of the data model are distinct (Little, 1995). This procedure can be implemented by using the following SAS (SAS, 2000) PROC MIXED code:

```
proc mixed method=ml;
class id group;
model score=time group group*time;
random intercept time/type=un subject=id;
```

The following are definitions of the variables used in this code:

- time—a quantitative index of the time of measurement
- id—a categorical variable identifying the participant
- group—a categorical variable identifying the treatment group

Pattern-Mixture Models (Unweighted Least Squares)

A number of different strategies have been presented over the years to deal with data that are MNAR [see the references provided by Little (1995) and Hedeker & Gibbons (1997)].

Recently, Little provided a general class of models referred to as pattern-mixture models. As Little (1995, p. 1113) noted, "Pattern-mixture models stratify the population by the pattern of dropout, implying a model for the whole population that is a mixture over the patterns." An advantage of this procedure is that the missing data mechanism is taken into account in the estimation, but a model for the pattern of missing data does not have to be explicitly introduced into the likelihood function.

A pattern-mixture model due to Little (1995), for the design considered in this paper, yields valid estimates of the treatment effect even when the pattern of missing data is related to the covariates and the subject specific slopes and intercepts (a type of MNAR missing data mechanism). The reader should note that Little (1995, p. 1120) indicated that the unweighted least squares (UWLS) estimate of the slope difference is the maximum likelihood (ML) estimator for the pattern-mixture model he used [see Equation (17) in Little] for analysis of longitudinal missing data under normal distribution theory. We implemented the UWLS procedure as follows:

1. Use ordinary least squares (OLS) to estimate the slope for each participant in each treatment group.

2. For each treatment calculate the unweighted average of the subject-specific OLS slopes,

$$\hat{\theta}_{1k} = \frac{\sum_{i=1}^{n_k} \hat{\beta}_{1ik}}{n_k},$$

and calculate the treatment effect as the difference between these two averages.

3. Calculate the sampling variance of the estimated treatment effect by using the (2,2) element of

$$S^2_{\hat{\theta}_{12}-\hat{\theta}_{11}} = \sum_{k=1}^{2} \frac{\sum_{i=1}^{n_k} \hat{\sigma}^2 \left(\mathbf{X}'_{ik}\mathbf{X}_{ik}\right)^{-1} + \hat{\mathbf{D}}}{n_k^2}$$

where the first column of $\mathbf{X}_{ik}$ is a vector of ones and the second column contains codes for the occasions on which participant *i* in group *k* had observed data. Wang-Clow et al. (1995) also used this method, however, they used the method of moments to calculate $\hat{\sigma}^2$ and $\hat{\mathbf{D}}$. We used ML estimation to calculate these quantities. Specifically, we used the PROC MIXED code used to implement the mixed model for MAR data. While these estimates assume that the missing data mechanism is not MNAR, comparison of the variance of $\hat{\theta}_{12} - \hat{\theta}_{11}$, over replications of a condition, to the average value of $S^2_{\hat{\theta}_{12}-\hat{\theta}_{11}}$ suggested that the method provides a consistent estimate of the sampling variance of $\hat{\theta}_{12} - \hat{\theta}_{11}$ for the conditions we studied.

Linear Minimum Variance Unbiased Estimator

Wu and Bailey (1989) presented a method which they called the linear minimum variance unbiased estimator. Later Wang-Clow et al. (1995) referred to the method as the ANCOVA method and we use the latter term in this paper. Wu and Bailey (1989) proposed using the following model within each group

$$\hat{\beta}_{1ik} = \gamma_{10k} + \gamma_{11} t_{ik} + \delta_{ik} \qquad (4)$$

where $\hat{\beta}_{1ik}$ is the OLS estimate of the subject-specific slope for participant *i* in group *k*. Wu and Bailey propose testing for a treatment effect by calculating an estimate of the expected value of β_{1ik}

$$\hat{E}\left(\beta_{1ik}\right) = \hat{\gamma}_{10k} + \hat{\gamma}_{11}\bar{t}_k, \qquad (5)$$

where $\bar{t}_k$ is the average in group *k* of t_{ik}, and comparing the estimates across treatment groups. Noting that the variance of $\hat{\beta}_{1ik}$ varies across treatment groups and the occasions on which the dependent variable was observed for participant *i*, Wu and Bailey proposed estimating the γs by weighted least squares (WLS) with the weight equal to the inverse of the sampling variance of $\hat{\beta}_{1ik}$. The sampling variance is the (2,2) element of $\hat{\sigma}^2 \left(\mathbf{X}'_{ik}\mathbf{X}_{ik}\right)^{-1} + \hat{\mathbf{D}}_k$. We implemented this WLS procedure. However, whereas Wu and Bailey and Wang-Clow et al. used method of moment estimators of $\hat{\sigma}^2$ and $\hat{\mathbf{D}}_k$, we used ML estimators obtained by using the following code:

```
proc mixed method=ml;
class id group;
model score=time group group*time/solution ;
random intercept time/type=un subject=id
group=group;
```

In the random statement the code group=group specifies that the covariance matrix for the intercept and slope varies across treatment groups.

The procedure described by Wu and Bailey (1989) is fairly complicated because of the necessity of estimating the weights used in the WLS procedure. However, Algina and Keselman (2003) reformulated the Wu and Bailey model as a multilevel model and estimated it by using PROC MIXED, thus eliminating the complication of estimating the weights. Their level-1 model is given by equation (1). The level 2 models are

$$\beta_{0ik} = \gamma_{00} + \gamma_{01} Z_{ik} + \gamma_{02}\left(t_{ik} - \bar{t}_k\right) + u_{0ik} \qquad (6)$$

and

$$\beta_{1ik} = \gamma_{10} + \gamma_{11} Z_{ik} + \gamma_{12}\left(t_{ik} - \bar{t}_k\right) + u_{1ik}. \qquad (7)$$

The estimate of the treatment effect is $\hat{\gamma}_{11}$ and testing $H_0 : \gamma_{11} = 0$ provides a test of the treatment effect. The approach presented by Wu and Bailey does not include an equation for the intercept. Nevertheless, Algina and Keselman included it because Bryk and Raudenbush (1992) have noted that omitting variables in one level-2 model can impact estimates in a second level-2 model because of the correlated error terms for the level-2 models. The model

represented by equations (1), (6), and (7) can be estimated by using the following PROC MIXED code:

```
proc mixed method=ml;
class id group;
model score=lobsc group time time*lobsc time*group /solution;
random intercept time/type=un subject=id group=group;
```

In the preceding code, the variable lobsc is $(t_{ik} - \bar{t}_k)$. The inclusion of lobsc and time*lobsc is intended to improve estimation and testing when the missing data mechanism is MNAR and the missing data pattern is monotone. If the data are MAR, it is known that valid estimates can be obtained with these terms excluded.

Analyses Investigated by Overall et al. (1999)

The simplest method studied by Overall et al. (1999) is an endpoint analysis. This analysis is a two-stage procedure. At stage one a simple change score from baseline to the last available measurement is calculated; at stage two the change scores are the dependent variable in an ANCOVA, using pretest score (Y_1) and time of the last observation as covariates and treatment group as the between-subjects factor.

Overall and his colleagues also used ANCOVA with PROC MIXED to examine the group by time effect (see Overall et al., 1999, pp. 205-209), using Y_1 and t_{ik} as covariates, though their approach differs from the Wu and Bailey (1989) approach. They use the following PROC MIXED code:

```
proc mixed;
class id group;
model score=lobs y1 group time time*group /solution;
random intercept time/type=un subject=id;
```

There are three major differences between the Overall et al. code and the PROC MIXED code used by Algina and Keselman (2003) to implement the Wu-Bailey procedure. First the time of last observation (lobs) is not centered. Second Y_1, the pretest score, is included in their model but not in the Algina-Keselman code. Third, the time by lobs interaction is excluded in their model. The result of this exclusion is that the time code for the last observation on which the participant was observed is excluded from the level-2 model for the slope. Thus, the Overall at al. PROC MIXED ANCOVA is based on the a multilevel model in which the level-1 model is given by equation (1) and the level-2 models are

$$\beta_{0ik} = \gamma_{00} + \gamma_{01} Z_{ik} + \gamma_{02} t_{ik} + \gamma_{03} Y_{1ik} + u_{0ik} \quad (8)$$

and

$$\beta_{1ik} = \gamma_{10} + \gamma_{11} Z_{ik} + \gamma_{12} t_{ik} + u_{1ik} . \quad (9)$$

The estimate of the treatment effect is $\hat{\gamma}_{11}$ and testing $H_0 : \gamma_{11} = 0$ provides a test of the treatment effect.

Overall et al. (1999) also investigated a two-stage ANCOVA procedure. Like the Wu and Bailey (1989) approach, Overall et al. use OLS in stage 1 to estimate the subject-specific regression coefficients and then these estimates, weighted by lobs, are used in a second stage ANCOVA with Y_1 and t_{ik} used as covariates.

Thus, the previously described analyses can be used to analyze the important group by time interaction effect in longitudinal designs in which data are missing. In this report we assess rates of Type I error and power in testing whether the average slopes are equal for the treatment groups, as well as the bias and variability (i.e., SD) in estimating the average slope difference.

Methodology

Algina and Keselman (2003) investigated three missing data mechanisms (CD, MAR and MNAR), but only considered monotone patterns. In the present investigation, whether or not data are missing for a participant is determined independently for each occasion. Therefore, the pattern of missing data is not monotone. In addition, eight different missing data mechanisms were used:

1. MNAR-Direct Selection (DS) on Y_j. The data point for participant *i* was missing at occasion *j* if $Y_{ij} > \delta$. The value of δ was selected so that the probability of missing data at time 3 was 5% for participants in treatment 2. This selection of δ determined the probability of missing data for both groups at time points 3 to 9. Figure 1 shows the probability of missing data at each occasion in treatments 1 and 2 in conditions in which there was a treatment effect. The δs for the other mechanisms were selected to yield the same probabilities of missing data. In conditions in which there was no treatment effect, the probability of missing data in treatment 1, at a particular occasion, was equal to the probability of missing data that is reported in Figure 1, at that occasion, for treatment 2.

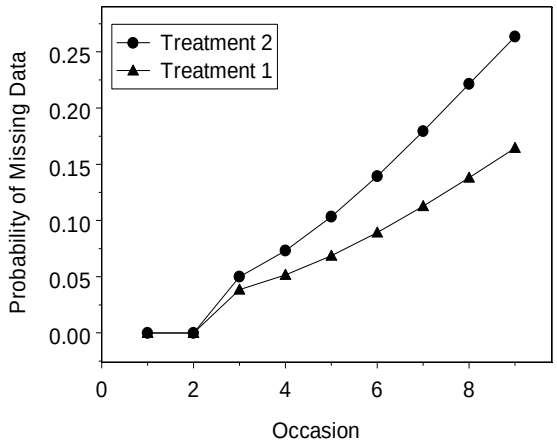

Figure1. Probability of Missing Data by Occasion.

2. CD. The data point for participant *i* was missing at occasion *j* if U_{ij} (a uniform random variable) was less than the probability determined for the MNAR condition with direct selection on Y.

3. MAR-DS. The criterion used to determine whether the data point for participant *i* was missing at occasion *j* depended on whether the data point for participant *i* was missing at occasion $j-1$: When the data point for participant *i* at occasion $j-1$ was not missing, the data point for participant *i* was missing at occasion *j* if $Y_{i(j-1)} > \delta_{j-1}$. When the data point for participant *i* at occasion $j-1$ was missing, the data point for participant *i* was missing at occasion *j* if U_{ij} was less than the probability determined for the MNAR condition with direct selection on Y.

If the first criterion had been used uniformly, the data would have been MNAR because, for a participant with missing data at occasion $j-1$, whether the data were missing at occasion *j* would depend on the value of a missing score at occasion $j-1$ rather than value of an observed score at occasion $j-1$.

4. MAR-Probabilistic Selection (PS). Again the criterion used to determine whether the data point for participant *i* was missing at occasion *j* depended on whether the data point for participant *i* was missing at occasion $j-1$: When the data point for participant *i* at occasion $j-1$ was not missing, the data point for participant *i* was missing at occasion *j* if $U_{ij} < \phi\left(\delta_{j-1} + Y_{i(j-1)}\right)$, where $\phi(\bullet)$ is the cumulative normal function. When the data point for participant *i* at occasion $j-1$ was missing, the data point for participant *i* was missing at occasion *j* if U_{ij} was less than the probability determined for the MNAR condition with direct selection on Y.

5. MNAR-DS on Y_{j-1}. The data point for participant *i* was missing at occasion *j* if $Y_{i(j-1)} > \delta_{j-1}$. This method employs the first criterion used in the MAR-DS mechanism.

6. MNAR-PS on Y_{j-1}. The data point for participant *i* was missing at occasion *j* if $U_{ij} < \phi\left(\delta_{j-1} + Y_{i(j-1)}\right)$. This method employs the first criterion used in the MAR-PS mechanism.

7. MNAR-PS on Y_j. The data point for participant *i* was missing at occasion *j* if $U_{ij} < \phi\left(\delta_j + Y_{ij}\right)$.

8. MNAR-PS on Slope and Intercept (SI). The data point for participant *i* was missing at occasion *j* if $U_{ij} < \phi\left(\delta_j + .46\beta_{0ik} + .14\beta_{1ik}\right)$.

The U_{ij} for the four probabilistic mechanisms were selected independently for each participant and time point.

It is impossible to know whether or not these eight missing data mechanisms are representative of those found in practice. However, these eight mechanisms represent a wider variety of mechanisms than have been included in previous research.

Seven methods of examining the group by time interaction effect in a randomized parallel groups design were examined; these methods were also examined by Algina and Keselman (2003). Specifically, the methods (with their acronyms) were:

1. Overall et al.'s (1999) two-stage endpoint ANCOVA (OEPAOC),
2. an unweighted least squares (pattern-mixture) analysis (UWLS),
3. the ANCOVA presented by Wang-Clow et al. (1995) (See Section 3.6 in their paper), where the weights for the WLS part of the analysis were obtained from PROC MIXED (WLSAOC),
4. Wu and Bailey's (1989) two-stage ANCOVA implemented in PROC MIXED (WBPMAOC),
5. Overall et al.'s (1999) PROC MIXED analysis that uses Y_1 and t_{ik} as covariates (OPMAOC),
6. Overall et al.'s (1999) two-stage ANCOVA (OTSAOC), and
7. The mixed model analysis, implemented in PROC MIXED, that presumes the data are missing at random (PMMAR).

Theory presented in Little (1995) shows that the UWLS estimator of the treatment effect is consistent when the data are CD or MNAR with missingness (i.e., whether a particular data point is missing) predicted by the slope and intercept. PMMAR is known to yield a consistent estimator when the data are CD or MAR. OEPAOC, WLSAOC, and WBPMAOC were designed to improve performance of the treatment effect estimator when the data are not MCAR or CD, but proofs of consistency have not been presented. Similarly, OPMAOC and OTSAOC were designed to improve control of the Type I error rate and power when the data are not MCAR or CD.

In addition to the eight types of missing data mechanism and the seven tests of the treatment effect, number of planned observations per group ($n_k = 100$ and $n_k = 200$) was also investigated. Overall and his colleagues (see Ahn, Tonidandel & Overall, 2000; Overall et al., 1999; Overall et al., 1996), as well as Algina and Keselman (2003), examined the group by time interaction effect in a parallel-groups design containing a baseline score and eight repeated measurements; thus, for comparative purposes we had nine levels for our number of repeated measurements.

To compare the procedures, we simulated data for a situation in which participants are randomly assigned to treatments. We used the following equation to generate data for the ith participant in group *k* on the *j*th occasion:

$$Y_{ijk} = \beta_{0ik} + \beta_{1ik} t_j + \varepsilon_{ijk}. \qquad (10)$$

The equation states that the data for the *i*th person on nine occasions has a linear relationship to the time of measurement. The *i* subscripts on the intercept (β_{0ik}) and slope (β_{1ik}) indicate that the intercept and slope vary across participants. We assumed

$$\begin{bmatrix} \beta_{0ik} \\ \beta_{1ik} \end{bmatrix} \sim N\left(\begin{bmatrix} \theta_{0k} \\ \theta_{1k} \end{bmatrix}, \mathbf{D} \right).$$

The mean for the intercept was 50 in both groups $(\theta_{01} = \theta_{02})$, implying that both treatment groups had the same population pretest mean. For Type I error data, the mean for the slope was 9.0 in treatments 1 and 2. That is, $\theta_{12} - \theta_{11} = 0$, indicating identical average rates of increase over time, hence a null condition. For our power comparisons, the mean for the slope was 4.5 in treatment 1 and 9.0 in treatment 2. Thus, $\theta_{12} - \theta_{11} = 4.5$. The errors ε_{ijk} were assumed to be uncorrelated for different times of observation. This does not imply that the scores were uncorrelated over time. Allowing the slope and intercept to vary across participants implies that scores were correlated over time. In all

cases the covariance matrix (**D**) for the intercept and slope was

$$\mathbf{D} = \begin{bmatrix} 15.21 & 12.42 \\ 12.42 & 82.81 \end{bmatrix}.$$

The correlation between the slope and intercept was .35, indicating that participants with higher pretest status increased more rapidly. The variance for the residuals, conditional on time was 240. Algina and Keselman (2003) also studied

$$\mathbf{D} = \begin{bmatrix} 15.21 & -12.42 \\ -12.42 & 82.81 \end{bmatrix},$$

but performance of WLSAOC and WBPMAOC was worse when $D_{12} > 0$, and so we have only included $D_{12} > 0$. The variable t_j is an index for observation time and was coded (0, 0.23077, 0.46154, 0.69231, 0.92308, 1.15385, 1.38462, 1.61538, 1.84615). The design of the simulation was based on Wang-Clow et al.'s (1995) study. In their study they had 14 time points, coded from 0 to 3. Our results would also have been obtained if we had coded t_j from 0 to 8 and had multiplied the β_{1ik} by $1.84615/8$.

Algina and Keselman (2003) also studied experiments with five time points. The performance of WLSAOC and WBPMAOC was worse with nine points and so we have elected to study only nine time points. Without further complications to the methods, the methods can only be applied to participants who have at least two observations. Therefore in our simulated data, every participant had an observation at the pretest and the first follow-up occasion. Each condition was replicated 1000 times. All hypothesis tests were conducted with a nominal alpha of .05.

Results

The slope difference $(\theta_{12} - \theta_{11})$ can be estimated by all procedures except OTSAOC and OEPAOC. For each condition in the study the slope difference was estimated by using each of the remaining five methods. Table 1 contains means and standard deviations of these estimates for the CD and MAR mechanisms. Comparison of the means to 0 when $\theta_{12} - \theta_{11} = 0$ and to 4.5 when $\theta_{12} - \theta_{11} = 4.5$ provides an indication of bias in the estimates. The standard deviations provide a measure of sampling variability of the estimates. The results indicate that all methods yielded unbiased estimators of the treatment effect when the missing data mechanism was CD and when the missing data mechanism was MAR and $\theta_{12} - \theta_{11} = 0$. However, when the missing data mechanism was MAR and $\theta_{12} - \theta_{11} = 4.5$ only PMMAR and OPMAOC yielded unbiased estimators. For a fixed sample size and a fixed value for the treatment effect there were no notable differences among the methods in the standard deviations of the estimates.

Table 2 contains estimated Type I error rates and power for the CD and MAR mechanisms. For CD data, all procedures had estimated Type I error rates near the nominal value and power differences were small but in favor of OEPAOC (Overall et al.'s, 1999 two-stage end-point procedure). For MAR data, WBPMAOC and WLSAOC had estimated Type I error rates above the nominal level. These two procedures are variations on the method suggested by Wu and Bailey (1989). For MAR data, OEPAOC and OTSAOC tended to have lower power than the other procedures. UWLS, WBPMAOC, and WLSAOC tended to have the best power, but this reflects the positively biased estimator produced by these three procedures. Comparing the two procedures that produced unbiased estimators of the treatment effect, PMMAR tended to have slightly better power than OPMAOC.

Tables 3 and 4 contain means and standard deviations of the estimated treatment effect for conditions in which the missing data mechanism was MNAR. Table 3 contains results for $\theta_{12} - \theta_{11} = 0$ and Table 4 contains results for $\theta_{12} - \theta_{11} = 4.5$. In both tables bold values indicate mean treatment effects that were significantly different from the population treatment effect. In Table 3, there was only one estimated treatment effect that was significantly different from 0 [$t(999) = 1.962$ for WBPMAOC and $n_k = 100$].

Table 1. Means and Empirical Standard Errors of Test Statistics for CD and MAR Conditions

		$\theta_{12}-\theta_{11}=0$						$\theta_{12}-\theta_{11}=4.5$					
		CD		MAR-DS		MAR-PS		CD		MAR-DS		MAR-PS	
n_k	Test	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
100	PMMAR	-0.014	1.863	-0.037	1.924	0.056	1.887	4.418	1.846	4.539	1.854	4.569	1.838
	UWLS	-0.013	1.881	-0.034	1.990	0.062	1.940	4.417	1.862	**4.858**	1.914	**4.878**	1.894
	OPMAOC	-0.016	1.863	-0.035	1.924	0.056	1.885	4.417	1.846	4.546	1.855	4.576	1.839
	WBPMAOC	-0.013	1.863	-0.046	1.995	0.052	1.948	4.420	1.851	**4.911**	1.915	**4.952**	1.881
	WLSAOC	-0.013	1.864	-0.044	1.997	0.055	1.954	4.417	1.852	**4.932**	1.921	**4.973**	1.890
200	PMMAR	-0.040	1.349	-0.024	1.296	0.043	1.284	4.501	1.251	4.451	1.310	4.492	1.303
	UWLS	-0.036	1.374	-0.028	1.354	0.049	1.327	4.505	1.259	**4.755**	1.357	**4.793**	1.353
	OPMAOC	-0.040	1.350	-0.025	1.296	0.044	1.284	4.501	1.251	4.457	1.310	4.496	.306
	WBPMAOC	-0.039	1.349	-0.026	1.359	0.054	1.330	4.503	1.251	**4.828**	1.355	**4.864**	1.347
	WLSAOC	-0.038	1.350	-0.027	1.364	0.054	1.335	4.503	1.250	**4.848**	1.358	**4.884**	1.353

Note: PMMAR-Proc Mixed MAR analysis; UWLS-Unweighted least squares analysis which is ML for pattern-mixture models; OPMAOC-Overall et al.'s (1999) Proc Mixed ANCOVA; WBPMAOC- Wu and Bailey's (1989) ANCOVA with PROC Mixed as defined in this paper; WLSAOC- Wang-Clow et al.'s (1995) ANCOVA analysis. Bold values indicate a statistically significant difference between the mean of $\widehat{\theta_{12}-\theta_{11}}$ and $\theta_{12}-\theta_{11}$.

Table 2. Estimated Type I Error Rates and Power

		CD		MAR-DS		MAR-PS	
n_k	Test	$\hat{\alpha}$	$1-\hat{\beta}$	$\hat{\alpha}$	$1-\hat{\beta}$	$\hat{\alpha}$	$1-\hat{\beta}$
100	PMMAR	0.044	0.670	0.053	0.685	0.048	0.711
	UWLS	0.048	0.661	0.063	0.738	0.057	0.745
	OPMAOC	0.039	0.646	0.044	0.664	0.038	0.687
	WBPMAOC	0.045	0.667	0.084	0.789	0.083	0.799
	WLSAOC	0.044	0.667	0.079	0.787	0.081	0.795
	OEPAOC	0.053	0.694	0.061	0.508	0.047	0.518
	OTSAOC	0.054	0.647	0.059	0.448	0.051	0.468
200	PMMAR	0.054	0.931	0.047	0.920	0.051	0.929
	UWLS	0.052	0.935	0.059	0.935	0.059	0.947
	OPMAOC	0.044	0.923	0.039	0.911	0.042	0.918
	WBPMAOC	0.053	0.930	0.086	0.955	0.076	0.972
	WLSAOC	0.052	0.929	0.082	0.956	0.076	0.971
	OEPAOC	0.044	0.950	0.047	0.797	0.047	0.780
	OTSAOC	0.058	0.919	0.040	0.742	0.043	0.745

Note: PMMAR-Proc Mixed MAR analysis; UWLS-Unweighted least squares analysis which is ML for pattern-mixture models; OPMAOC-Overall et al.'s (1999) Proc Mixed ANCOVA; WBPMAOC- Wu and Bailey's (1989) ANCOVA with PROC Mixed as defined in this paper; WLSAOC- Wang-Clow et al.'s (1995) ANCOVA analysis; WLSAOCMM-Wang-Clow et al.'s ANCOVA using the method of moments for estimation; OEPAOC- Overall et al.'s two-stage endpoint ANCOVA analysis; OTSAOC-Overall et al.'s two-stage ANCOVA.

Table 3 Means and Estimated Standard Errors of Test Statistics: MNAR and $\theta_{12} - \theta_{11} = 0$

		MNAR-DS-Y_{j-1}		MNAR-PS-Y_{j-1}		MNAR-DS-Y_j		MNAR-PS-Y_j		MNAR-PS-SI	
n_k	Test	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
100	PMMAR	0.068	1.825	0.041	1.802	0.006	1.454	-0.086	1.524	-0.041	1.754
	UWLS	0.077	2.064	0.041	2.090	0.062	1.561	-0.094	1.616	-0.070	3.110
	OPMAOC	0.068	1.827	0.044	1.810	0.008	1.454	-0.086	1.522	-0.047	1.764
	WBPMAOC	0.072	2.039	0.060	2.065	0.012	1.525	**-0.098**	1.572	-0.071	2.084
	WLSAOC	0.073	2.045	0.064	2.078	0.013	1.519	-0.092	1.568	-0.073	2.019
200	PMMAR	-0.045	1.274	0.044	1.251	0.040	1.077	-0.013	1.048	0.012	1.284
	UWLS	-0.065	1.495	0.045	1.455	0.041	1.193	-0.023	1.146	0.009	2.165
	OPMAOC	-0.045	1.278	0.043	1.258	0.044	1.080	-0.015	1.046	0.019	1.291
	WBPMAOC	-0.050	1.460	0.027	1.432	0.048	1.130	-0.016	1.076	0.062	1.475
	WLSAOC	-0.051	1.468	0.030	1.437	0.049	1.128	-0.016	1.077	0.055	1.447

Note: See note to Table 1.

Table 4. Means and Estimated Standard Errors of Test Statistics: MNAR and $\theta_{12} - \theta_{11} = 4.5$

		MNAR-DS-Y_{j-1}		MNAR-PS-Y_{j-1}		MNAR-DS-Y_j		MNAR-PS-Y_j		MNAR-PS-SI	
n_k	Test	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
100	PMMAR	**4.314**	1.875	**4.287**	1.781	**2.937**	1.540	**2.833**	1.493	**3.859**	1.809
	UWLS	**4.990**	2.073	**4.978**	1.998	**3.218**	1.667	**3.141**	1.629	4.596	3.048
	OPMAOC	**4.364**	1.880	**4.336**	1.785	**3.026**	1.542	**2.921**	1.494	**4.024**	1.819
	WBPMAOC	**5.165**	2.051	**5.132**	1.990	**3.419**	1.589	**3.310**	1.552	**4.992**	2.077
	WLSAOC	**5.182**	2.059	**5.149**	2.000	**3.396**	1.585	**3.289**	1.552	**4.845**	2.045
200	PMMAR	**4.305**	1.269	**4.294**	1.328	**2.879**	1.007	**2.873**	1.073	**3.815**	1.218
	UWLS	**4.967**	1.417	**4.973**	1.477	**3.168**	1.082	**3.149**	1.140	4.457	2.062
	OPMAOC	**4.351**	1.272	**4.342**	1.333	**2.970**	1.005	**2.961**	1.067	**3.988**	1.236
	WBPMAOC	**5.140**	1.394	**5.128**	1.455	**3.366**	1.037	**3.340**	1.079	**4.933**	1.458
	WLSAOC	**5.158**	1.404	**5.151**	1.468	**3.339**	1.036	**3.319**	1.079	**4.794**	1.425

Note: See note to Table 1.

Table 5. Estimated Type I Error Rates and Power: MNAR Conditions.

		MNAR-DS-Y_{j-1}		MNAR-PS-Y_{j-1}		MNAR-DS-Y_j		MNAR-PS-Y_j		MNAR-PS-SI	
n_k	Test	$\hat{\alpha}$	$1-\hat{\beta}$	$\hat{\alpha}$	$1-\hat{\beta}$	$\hat{\alpha}$	$1-\hat{\beta}$	$\hat{\alpha}$	$1-\hat{\beta}$	$\hat{\alpha}$	$1-\hat{\beta}$
100	PMMAR	0.041	0.661	0.051	0.665	0.058	0.509	0.066	0.477	0.045	0.586
	UWLS	0.058	0.731	0.072	0.738	0.043	0.530	0.050	0.503	0.043	0.362
	OPMAOC	0.034	0.633	0.041	0.641	0.049	0.498	0.060	0.465	0.033	0.580
	WBPMAOC	**0.111**	0.825	**0.116**	0.825	0.064	0.632	**0.083**	0.607	**0.092**	0.756
	WLSAOC	**0.108**	0.822	**0.117**	0.822	0.065	0.625	**0.079**	0.605	0.066	0.730
	OEPAOC	0.048	0.500	0.047	0.498	0.048	0.484	0.042	0.427	0.037	0.521
	OTSAOC	0.043	0.444	0.051	0.457	0.051	0.418	0.055	0.404	0.048	0.478
200	PMMAR	0.054	0.924	0.048	0.906	0.067	0.762	0.044	0.769	0.039	0.849
	UWLS	0.075	0.948	0.067	0.950	0.053	0.780	0.040	0.775	0.050	0.589
	OPMAOC	0.043	0.912	0.036	0.893	0.050	0.764	0.035	0.775	0.033	0.863
	WBPMAOC	**0.115**	0.976	**0.111**	0.978	**0.076**	0.876	0.056	0.886	**0.092**	0.957
	WLSAOC	**0.111**	0.974	**0.114**	0.977	**0.077**	0.872	0.059	0.880	**0.079**	0.948
	OEPAOC	0.042	0.792	0.048	0.797	0.058	0.737	0.051	0.732	0.037	0.836
	OTSAOC	0.051	0.742	0.036	0.741	0.053	0.680	0.046	0.677	0.055	0.749

Note. See note to Table 2. Bold values indicate $\hat{\alpha} > .075$

UWLS, WBPMAOC, and WLSAOC tended to have slightly larger standard deviations than did PMMAR and OPMAOC. In Table 4 all treatment effects were significantly different from 4.5 except for UWLS under the MNAR-PS-SI condition. Again WBPMAOC and WLSAOC tended to have slightly larger standard deviations than did PMMAR and OPMAOC. Except in the MNAR-PS-SI conditions, UWLS tended to have standard deviations similar to those for WBPMAOC and WLSAOC. In the MNAR-PS-SI conditions UWLS had notably larger standard deviations than did the other procedures.

Table 5 contains estimated Type I error rates and power for MNAR missing data mechanisms. With regard to Type I error control we note that, as was true with MAR data, WBPMAOC and WLSAOC did not control the Type I error rate.

Of the methods that control their rates of Type I error, the methods divide into two groups: the more powerful methods (PMMAR, UWLS, and OPMAOC) and the less powerful methods (OEPAOC and OTSAOC). The difference in power between the two groups was quite substantial in most conditions. When missingness was predicted by slopes and

intercepts (MNAR-PS-SI), PMMAR and OPMAOC were more powerful than UWLS. In the other MNAR conditions, UWLS was more powerful than PMMAR or OPMAOC. The power advantage in favor of UWLS was smaller than the power advantage for PMMAR and OPMAOC.

Conclusion

Presented and examined are methods of analysis that, according to the literature, should result in better estimation of unknown parameters and which take MNAR missingness into account in their analyses when data are missing in a parallel groups design. In particular, we investigated methods due to Little (1995), Wu and Bailey (1988, 1989), Wang-Clow et al. (1995) and Overall et al. (1999).

The results, along with those in Algina and Keselman (2003), suggest that whether the pattern of missing data is monotone or not will influence the selection of a method for analyzing the data. Based on bias, control of Type I error rate, and power, Algina and Keselman concluded that Overall at al.'s (1998) mixed model procedure (OPMAOC) is promising when the missing data pattern is monotone.

The present research suggests that OPMAOC works reasonably well when the missing data pattern is not monotone, but that the mixed model for MAR data (PMMAR) and UWLS are very competitive. Comparing OPMAOC and PMMAR, both controlled the Type I error rate in all conditions investigated and power differences were very small. The major difference was that under the MNAR missing data mechanism OPMAOC tended to be slightly less biased than PMMAR was. Comparing OPMAOC and PMMAR, both controlled the Type I error rate. Power differences depended on the missing data mechanism.

When missingness was predicted by the slope and intercept, the missing data mechanism for which UWLS was developed, UWLS was much less powerful then OPMAOC because its standard error was notably larger than the standard error for OPMAOC. In all other conditions, power for the two procedures was either quite similar or favored UWLS. Bias differences also depended on the missing data mechanism. When missingness was predicted by the slope and intercept, UWLS was unbiased but OPMAOC was not. When the data were MAR, OPMAOC was unbiased but UWLS was not. When missingness on Y_j was predicted by scores on Y_j (MNAR-DS-Y_j and MNAR-PS-Y_j), UWLS was less biased than was OPMAOC. The opposite was true when missingness was predicted by scores on Y_{j-1} (MNAR-DS-Y_{j-1} and MNAR-PS-Y_{j-1}).

Considering the performance of OPMAOC in Algina and Keselman (2003) and in the present study, if a researcher wants to use a single procedure for monotone and non-monotone patterns of missing data, OPMAOC appears promising. Of course, as is true of all empirical studies, the generalizability of the results is limited by the design of the study. The procedures may perform differently if different models for dropping out are adopted.

References

Algina, J., & Keselman, H. J. (2003). Analyzing effects in longitudinal studies with missing data. *Journal of Modern Applied Statistical Methods, 2*, 50-72.

Ahn, C., Tonidandel, S., & Overall, J. E. (2000). Issues in use of SAS Proc Mixed to test the significance of treatment effects in controlled clinical trials. *Journal of Biopharmaceutical Statistics, 10(2)*, 265-286.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Diggle, P. J., & Kenward, M. G. (1994). Informative dropout in longitudinal data analysis, with discussion. *Applied Statistics, 43* 49-73.

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*, 64-78.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112-1121.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Overall, J. E., Ahn, C., Shivakumar, C., & Kalburgi, Y. (1999). Problematic formulations of SAS Proc Mixed models for repeated measurements. *Journal of Biopharmaceutical Statistics*, *9*, 189-216.

Overall, J. E., Ghasser, S., & Fiore, J. (1996). Random regression with imputed values for dropouts. *Psychopharmacology Bulletin*, *32*, 377-388.

SAS Institute Inc. (2000). *SAS/STAT user's guide, version 8*. Cary, NC: Author.

Wang-Clow, F., Lange, M., Laird, N. M., & Ware, J. H. (1995). A simulation study of estimators for rates of change in longitudinal studies with attrition. *Statistics in Medicine*, *14*, 283-297.

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

Verbeke, G., & Molenberghs. (2000). *Linear mixed models for longitudinal data*. New York: Springer.