Journal of Modern Applied Statistical Methods

Volume 2 | Issue 1 Article 6

5-1-2003

Analyzing Group by Time Effects in Longitudinal Two-Group Randomized Trial Designs With Missing Data

James Algina
University of Florida, algina@ufl.edu

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

Abdul R. Othman
Universiti Sains, Malaysia, oarahman@usm.my

Journal of Modern Applied Statistical Methods
May, 2003, Vol. 2, No. 1, 50-72

1538 – 9472/03/$30.00

Analyzing Group by Time Effects in Longitudinal Two-Group Randomized Trial Designs With Missing Data

James Algina
University of Florida

H. J. Keselman
University of Manitoba

A. R. Othman
Universiti Sains Malaysia

We investigated bias, sampling variability, Type I error and power of nine approaches for testing the group by time interaction in a repeated measures design under three types of missing data mechanisms. One procedure due to Overall, Ahn, Shivakumar, and Kalburgi (1999) performed reasonably well over a range of conditions.

Key words: Missing data, random coefficients model, pattern mixture model

Introduction

Consider a design in which N participants are randomly assigned to $K = 2$ treatments. The researcher plans to observe each participant J times on the dependent variable, with the first observation prior to initiating a treatment and the remaining $J-1$ observations following initiation of a treatment.

This design has been referred to as a longitudinal two-group randomized trial design (Delucchi & Bostrom, 1999), randomized parallel-groups design (Overall, Ghasser, Shobaki & Fiore, 1996), or split-plot repeated measures design (Littell, Milligan, Stroup, & Wolfinger, 1996; Maxwell & Delaney, 1990). The effect of primary interest, typically, is whether there are differential rates of change over time, that is, whether there is a group by time interaction.

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

Let Y_{ijk} denote a random variable underlying the score, in treatment k $(k = 1,2)$, for participant i $(i = 1,\cdots,n_k)$, on occasion j $(j = 1,\cdots,J)$. A possible model for the subject-specific regression of the dependent variable on time of measurement is

$$\mathbf{y}_{ik} = \mathbf{X}\mathbf{b}_{ik} + \mathbf{e}_{ik}$$

where $\mathbf{y}'_{ik} = (Y_{i1k},\cdots,Y_{iJk})$, $\mathbf{b}_{ik}$ is an unobservable r-dimensional random vector, $\mathbf{e}_{ik}$ is a J-dimensional random vector,

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{r-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_J & t_J^2 & \cdots & t_J^{r-1} \end{bmatrix},$$

and $t_1,\cdots,t_J$ indexes time of measurement. We assume $\mathbf{e}_{ik} \sim N(0, s^2\mathbf{I}_J)$.

In this paper we focus on situations in which it is reasonable to assume that the subject-specific regressions are well described by a linear trend. Therefore

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_J \end{bmatrix}$$

and $b'_{ik} = (b_{0ik}\ b_{1ik})$. The between-subjects model for b_{ik} is

$$\begin{bmatrix} b_{0ik} \\ b_{1ik} \end{bmatrix} = \begin{bmatrix} 1 & z & 0 & 0 \\ 0 & 0 & 1 & z \end{bmatrix} \begin{bmatrix} g_{00} \\ g_{01} \\ g_{10} \\ g_{11} \end{bmatrix} + \begin{bmatrix} u_o \\ u_1 \end{bmatrix} \qquad (1)$$

where $z = 0$ for the first treatment and 1 for the second treatment. More compactly $b_{ik} = \mathbf{W}g + \mathbf{u}$. We assume that $\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$.

In many studies, participants may not be observed on all occasions. In general, the correct method of analysis depends on the missing data mechanism. Using an incorrect method can result in inconsistent estimates of the parameters. Little (1995) reviewed two different classes of methods for use in longitudinal designs. The design considered in this paper is a special case of the longitudinal design considered by Little. Little presented his review in the context of monotone missing data patterns, a context we adopt here. That is, we assume that if a participant is not observed on a particular occasion, the participant is not observed on any subsequent occasion.

Random Coefficient Models

Let J_{ik} denote the last occasion at which participant i in group k was observed and $t_{J_{ik}}$ the value of *t* for this time point and $\mathbf{y}_{ik}$ be partitioned as $\mathbf{y}'_{ik} = (\mathbf{y}'_{obs,\,ik}\ \mathbf{y}'_{miss,\,ik})$, $R_{ik} = J$ if the participant has complete data, and $R_{ik} = J_{ik}$, otherwise. The first class of methods is the random coefficient selection models. According to Little (1995), in this approach the joint distribution of $\mathbf{y}_{ik}$, b_{ik}, and R_{ik} is factored as

$$f(\mathbf{y}_{ik}, b_{ik}, R_{ik} \mid \mathbf{X}, \mathbf{W}) = f(\mathbf{y}_{ik} \mid \mathbf{X}, \mathbf{W}, b_{ik})\, f(b_{ik} \mid \mathbf{W})\, f(R_{ik} \mid \mathbf{X}, \mathbf{W}, \mathbf{y}_{ik}, b_{ik}).$$

In our context, the model for $f(\mathbf{y}_{ik} \mid \mathbf{X}, \mathbf{W}, b_{ik})$ is

$$(\mathbf{y}_{ik} \mid \mathbf{X}, \mathbf{W}, b_{ik}) \sim N(\mathbf{W}g + \mathbf{X}\mathbf{u}, s^2\mathbf{I}_J)$$

and

$$(b_{ik} \mid \mathbf{W}) = \mathbf{u} \sim N(\mathbf{0}, \mathbf{D}).$$

The model for $f(R_{ik} \mid \mathbf{X}, \mathbf{W}, \mathbf{y}_{ik}, b_{ik})$ is the model for the missing data mechanism. The data are referred to as missing completely at random (MCAR) if

$$f(R_{ik} \mid \mathbf{X}, \mathbf{W}, \mathbf{y}_{ik}, b_{ik}) = f(R_{ik})$$

(see Rubin, 1976; Little, 1995; Little & Rubin, 1987). That is, the data are MCAR if the probability of a particular data point being missing does not depend on either $\mathbf{y}_{ik}$, b_{ik}, $\mathbf{X}$ or $\mathbf{W}$. The missing data mechanism is called missing at random (MAR) if

$$f(R_{ik} \mid \mathbf{X}, \mathbf{W}, \mathbf{y}_{obs,\,ik}, \mathbf{y}_{miss,\,ik}, b_{ik}) = f(R_{ik} \mid \mathbf{X}, \mathbf{W}, \mathbf{y}_{obs,\,ik}),$$

that is, the probability of a particular data point being missing does not depend on either $\mathbf{y}_{miss,ik}$ or b_{ik}. Following Verbeke and Molenberghs (2000, p. 213), a missing data mechanism that does not meet either of these criteria can be referred to as missing not at random (MNAR). Consistent estimates for g can be obtained from the likelihood for $\mathbf{y}_{obs,\,ik}$ and R_{ik}. However if the data are MCAR or MAR (and if the parameters of the missing data mechanism are distinct from the parameters for the data), consistent estimates can be obtained by maximizing the likelihood for $\mathbf{y}_{obs,\,ik}$, a process that is called ignoring the missing data mechanism. Thus, for the purposes of estimating the fixed effects, the missing data mechanism is ignorable if the mechanism is MCAR or MAR, but the missing data mechanism is non-ignorable if the mechanism is MNAR.

As Hedeker and Gibbons (1997) noted "many instances of missing data are related to previous performance or other subject characteristics..." [See Little (1995, Section 2.2.2) and Schafer (1997, Ch. 2) for other examples of studies where MAR is a reasonable model of missingness]. Accordingly, MAR may very well be a reasonable process to presume for the missing data in one's study. Again, it should be noted for completeness, that in order to legitimately ignore the missing data mechanism for estimation

random but, as well, the parameters of the missing data mechanism must be independent of the parameters of the data model (Little, 1995; Little & Rubin, 1987; Schafer, 1997). This independence or distinctness of parameters is quite realistic in many contexts (See Schafer, 1997, pp. 11-15). When the missing data mechanism is ignorable, numerical results can easily be obtained with commercially available software, e.g., the SAS PROC (SAS, 1995) MIXED program (See Littell et al., 1996).

Pattern Mixture Models

The second class of models presented by Little (1995) is the class of random coefficient 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 when drop-out depends on $\mathbf{X}$, $\mathbf{W}$ and b_{ik} but not on $\mathbf{y}_{ik}$, the missing data mechanism does not have to be explicitly introduced into the likelihood function.

According to Little (1995), pattern-mixture models are based on the factorization

$$f(\mathbf{y}_{ik}, b_{ik}, R_{ik} \mid \mathbf{X}, \mathbf{W}) = f(\mathbf{y}_{ik} \mid \mathbf{X}, \mathbf{W}, b_{ik}, R_{ik})\, f(b_{ik} \mid \mathbf{W}, R_{ik})\, f(R_{ik} \mid \mathbf{W}).$$

In this expression $f(\mathbf{y}_{ik} \mid \mathbf{X}, \mathbf{W}, b_{ik}, R_{ik})$ models the subject-specific regressions stratified by missing data pattern, $f(b_{ik} \mid \mathbf{W}, R_{ik})$ models the subject-specific regression coefficients as a function of the between-subjects variables and the missing-data pattern, and $f(R_{ik} \mid \mathbf{W})$ models the proportions of each missing data pattern as functions of the between-subjects variables. The approach stratifies the sample by time and missing data pattern and models differences in the distributions of the dependent variables over these patterns.

Little (1995, p. 1118) presented a pattern-mixture model in which $e_{ik} \sim N(\mathbf{0}, s^2\mathbf{I}_J)$, as in the model considered in this paper, and drop-out depends on $\mathbf{W}$ and b_{ik} but not on $\mathbf{y}_{ik}$. In this case

$$(\mathbf{y}_{ik} \mid \mathbf{X}, \mathbf{W}, b_{ik}, R_{ik} = J_{ik}) \sim N(\mathbf{W}g^{(j)} + \mathbf{X}\mathbf{u}, s^2\mathbf{I}_J) \quad (2)$$

and

$$(b_{ik} \mid \mathbf{W}) = \mathbf{u} \sim N(\mathbf{0}, \mathbf{D}). \quad (3)$$

The notation $g^{(j)}$ indicates that the fixed effects introduced in equation (1) depend on drop-out time. Let p_{jk} denote the probability that a participant in treatment k drops out after occasion j. The pattern-mixture model estimate of the treatment effect is

$$\sum_j \hat{p}_{j2}\left(\hat{g}_{10}^{(j)} + \hat{g}_{11}^{(j)}\right) - \sum_j \hat{p}_{j1}\hat{g}_{10}^{(j)}. \quad (4)$$

Little pointed out that the $g^{(j)}$ can be estimated in PROC MIXED by introducing drop-out time as a categorical variable. The standard error can be computed using the delta method.

Another alternative is to use the un-weighted least squares (UWLS) approach presented by Wang-Clow, Lange, Laird, and Ware (1995). As Little (1995, p. 1120) noted, UWLS is maximum likelihood for the pattern-mixture model described in equations (2) and (3). In the UWLS approach, the estimated treatment effect is

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\hat{b}_{1i1} - \frac{1}{n_2}\sum_{i=1}^{n_2}\hat{b}_{1i2} \quad (5)$$

where $\hat{b}_{1ik}$ is the ordinary least squares (OLS) estimate of the subject-specific slope for the ith subject in the kth group. The standard error of the estimated treatment effect is the (2,2) element of

$$\sum_k \sum_i \frac{\widehat{\mathbf{V}}_i}{n_k^2} \quad (6)$$

where $\widehat{\mathbf{V}}_i = \hat{s}^2(\mathbf{X}_i'\mathbf{X}_i)^{-1} + \hat{\mathbf{D}}$ and

$$\mathbf{X}_i = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_{J_i} \end{bmatrix}.$$

Wang-Clow et al. (1995) showed how to estimate $\hat{s}^2$ and $\hat{\mathbf{D}}$ using the method of moments. These

quantities can also be estimated by using maximum likelihood.

Pattern-mixture modeling is potentially an important approach to analyzing longitudinal data collected in the design considered in this study. However, the method does have one drawback. The results of simulation studies reported by Wu and Carroll (1988), Wu and Bailey (1989), and Wang-Clow et al. (1995) indicated that when the pattern-mixture model in equations (2) and (3) is used the maximum likelihood estimate of the treatment effect may be highly inefficient. For example, Wang-Clow et al. compared various estimation procedures [e.g., un-weighted least squares, maximum likelihood, generalized least squares) under a number of missing data mechanisms (e.g., MAR and MNAR) in a two-group longitudinal design in which measurements were taken over 14 occasions. Wang-Clow et al. tabulated the sampling mean and standard deviation (sd) of the estimated treatment difference between mean slopes (see their Table II), and Type I error and power rates for the test of the treatment difference between mean slopes (see their Table III).

The treatment difference between mean slopes estimates the treatment effect. With regard to their Table II results, the sds for the UWLS method were frequently considerably larger than the other estimation procedures (e.g., under one of their MNAR cases, the UWLS sd was 41.62, while the values for the other estimators ranged from 16.97 to 18.05). The MSE for the UWLS estimator, again under one of the MNAR mechanisms, was 1730.80, a value much larger than those reported for the other estimators (range = 320.51-562.47).

Consequently, Wang-Clow et al. in their summary indicated that "the unweighted estimator is too inefficient to merit consideration." (p. 294). (Of course, this conclusion may be limited to the conditions of their simulation.) They drew this conclusion despite the fact that the pattern-mixture model estimator of the treatment effect was unbiased in all conditions. Finally, Type I error rates were frequently conservative (range 3.2%-3.8%) and importantly, power to detect differences was considerably less than when other estimators were used (e.g., 15.3% vs. 10.5%-32%).

Hedeker and Gibbons (1997) presented an example illustrating application of the pattern-mixture model approach to data collected in the design considered in this paper. Whereas Little's (1995) presentation indicated stratifying participants into as many strata as there are missing data patterns, Hedeker and Gibbons argued that, when the number of participants in some of the strata is small, the strata containing these participants can be combined. In their example, Hedeker and Gibbons had two strata. One included all participants who had a measurement on the last measurement occasion; the other included all other participants. Both groups included participants with different missing data patterns.

The potential problem with this approach can be seen by contrasting it with the UWLS approach used by Wang-Clow et al. (1995). Recall that this approach is maximum likelihood for the pattern-mixture model described in equations (2) and (3). In UWLS, the OLS estimate of the subject-specific slope is calculated for each participant. The un-weighted average of these slopes is then computed for each treatment group and the estimated treatment effect is the difference between these averages. The same estimate would be obtained if participants were stratified into as many strata as there are missing data patterns and ML were applied. This follows because the ML estimate of the expected value of b_{ik} within stratum j and treatment group k is

$$\hat{B}_{kj} = \frac{\sum_{i=1}^{n_{kj}} \hat{\mathbf{V}}_i^{-1} \hat{b}_{ik}}{\sum_{i=1}^{n_{kj}} \hat{\mathbf{V}}_i^{-1}},$$

where $\hat{b}_{ik}$ is the OLS estimate of b_{ik}. When there are as many strata as missing data patterns, within a stratum and treatment group $\hat{\mathbf{V}}_i$ is a constant over i and $\hat{B}_{kj}$ is the un-weighted average of the OLS estimates. Then, the estimated treatment effect is the second element of

$$\sum_j \hat{p}_{j2} \hat{B}_{2j} - \sum_j \hat{p}_{j1} \hat{B}_{1j},$$

which is equivalent to equations (4) and (5). On the other hand, when the strata are combined as

suggested by Hedeker and Gibbons, the $\hat{\mathbf{V}}_i$ are not constant over i and the ML estimate of the expected value of b_{ik} within a stratum and treatment is a weighted average of the least squares estimates of the subject-specific slopes for that group. Then, if the expected values of the within-subject regression parameters vary over the missing data patterns that were combined into the missing-data groups, the Hedeker-Gibbons' approach, with two strata, to the pattern-mixture model is likely to yield inconsistent estimators even when the missing data conform to the missing data mechanism assumed by the model in equations (2) and (3).

The Hedeker and Gibbons (1997) model is

$$Y_{ijk} = b_{0ik} + b_{1ik}t_j + e_{ijk} \quad (7)$$

$$b_{0ik} = l_{00} + l_{01}z + l_{02}z_2 + l_{03}(z \times z_2) + u_{0i} \quad (8)$$

$$b_{1ik} = l_{10} + l_{11}z + l_{12}z_2 + l_{13}(z \times z_2) + u_{1i} \quad (9)$$

where z_2 is 0 for participants with complete data and 1 otherwise. Using the gamma coefficients defined in equation (1), this model can also be written explicitly as a pattern-mixture model

$$\begin{aligned} Y_{ijk} = {} & g_{00}^{(z_2)} + g_{01}^{(z_2)}z + g_{10}^{(z_2)}t_j \\ & + g_{12}^{(z_2)}(z \times t_j) + u_{1j}t_j + u_{0j} + e_{ijk} \end{aligned} \quad (10)$$

where, as in equation (2), the superscript indicates the group (drop-out or completer) which the parameter describes. Using this notation $g_{12}^{(0)}$ is the treatment effect for the completers (i.e., the Time $\times$ Treatment interaction for the completers) and $g_{12}^{(1)}$ is the treatment effect for the dropouts. Further, $\hat{l}_{11}$ estimates $g_{12}^{(0)}$ and $\hat{l}_{13}$ estimates $g_{12}^{(1)} - g_{12}^{(0)}$ (the difference in the Time $\times$ Treatment interaction for the drop-outs and completers). Therefore the estimated treatment effect is $\hat{p}_c\hat{l}_{11} + \hat{p}_d(\hat{l}_{11} + \hat{l}_{13})$ where $\hat{p}_c$ and $\hat{p}_d$ are the estimated proportion of participants who completed and dropped out, respectively. The estimated sampling variance is

$$\hat{p}_c^2 V(\hat{l}_{11}) + \hat{p}_d^2 \times V(\hat{l}_{11} + \hat{l}_{13}) + \frac{\hat{p}_c\hat{p}_d \times \hat{l}_{13}^2}{n_1 + n_2}$$

where

$$V(\hat{l}_{11} + \hat{l}_{13}) = V(\hat{l}_{11}) + V(\hat{l}_{13}) + 2C(\hat{l}_{11}, \hat{l}_{13}),$$

$V(\bullet)$ denotes a sampling variance and $C(\bullet,\bullet)$ denotes a sampling covariance.

Alternative Methods

A number of other analytic methods, that use information about the pattern of missing data, have been suggested in the literature and one of our goals in this paper is to review alternative methods for analyzing effects in longitudinal designs in which data are missing; the second goal is to report the results of a simulation study which compares the methods.

Wu and Bailey (1989) presented an alternative 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. Provided participants are randomly assigned to groups and it is reasonable to assume that the subject-specific regressions of the dependent variable on time of measurement are 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. Wu and Bailey proposed the following procedure:

1. Use OLS to estimate the slope for each participant in each treatment group.

2. Using the estimated slopes as the dependent variable, conduct an ANCOVA with treatment group as the between-subjects factor of interest. Wu and Bailey discussed including two types of covariates. The first is the time point after which the participant dropped out and the second comprises the pretreatment score on the variable of interest and other pretreatment measures that may be available. In this paper we investigate the model without the second type of covariate, as did Wu and Bailey and Wang-Clow et. al (1995). However, we also investigate a related procedure due to Overall, Ahn, Shivakumar, and Kalburgi (1999) that includes the pretest as the covariate.

Wu and Bailey showed that the error variance in this model will vary over dropout times and presented a weighted least squares procedure for estimation and hypothesis testing. The test for the treatment effect (i.e., the group $\times$ time interaction) is the test of the treatment factor in the ANCOVA. In calculating the weights, Wu and Bailey assumed

$$b_{ik} \sim N(\mathbf{B}_k, \mathbf{D}_k).$$

Wu and Bailey presented method of moment estimators for $\mathbf{D}_k$ and s^2. Alternatively, maximum likelihood estimates for $\mathbf{D}_k$ and s^2 can be obtained by using PROC MIXED:

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

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

- id-a categorical variable identifying the participant
- group-a categorical variable identifying the treatment 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 to implement because of the necessity of estimating the weights and inserting them in a weighted least squares procedure. However, we show that a related procedure can be easily implemented in PROC MIXED. Wu and Bailey proposed using the following model to compare treatment groups:

$$\hat{b}_{1ik} = l_{10k} + l_{11} t_{J_{ik}} + d_{ik}.$$

They compare the groups by using

$$\hat{l}_{10k} + \hat{l}_{11}\bar{t}_k,$$

where $\bar{t}_k$ is the average of $t_{J_{ik}}$ for the kth group. If the model

$$\hat{b}_{1ik} = l_{10} + l_{11}\left(t_{J_{ik}} - \bar{t}_k\right) + l_{12} z + d_{ik} \qquad (11)$$

is estimated, then

$$\hat{l}_{12} = \left(\hat{l}_{102} - \hat{l}_{101}\right) + \hat{l}_{11}\left(\bar{t}_2 - \bar{t}_1\right).$$

An alternative to equation (11) is

$$b_{1ik} = l_{10} + l_{11}\left(t_{J_{ik}} - \bar{t}_k\right) + l_{12} z + u_{i1}. \qquad (12)$$

Readers familiar with multilevel models will recognize this model as a level-2 model for the slope in the level-1 equation

$$Y_{ijk} = b_{0ik} + b_{1ik} t_j + e_{ijk}. \qquad (13)$$

We also formulate a level-2 model for the intercept:

$$b_{0ik} = l_{00} + l_{01}\left(t_{J_{ik}} - \bar{t}_k\right) + l_{02} z + u_{i0}. \qquad (14)$$

The approach presented by Wu and Bailey (1989) does not include an equation for the intercept. Nevertheless, we include it because Bryk and Raudenbush (1992) have noted that omitting variables in one level-2 model can impact estimates in a second equation because of the correlated error terms for the level-2 models. By including $\left(t_{J_{ik}} - \bar{t}_k\right)$ in equations (12) and (14), the model conditions on the missing data pattern and the model can be formulated as a pattern-mixture model.

PROC MIXED can estimate the model represented by equations (12) to (14). The PROC MIXED program we suggest using is:

```
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;
```

The variable lobsc is $\left(t_{J_{ik}} - \bar{t}_k\right)$. The inclusion of lobsc and time*lobsc is intended to improve estimation and testing when drop-out depends on $\mathbf{W}$ and b_{ik} as in Little's (1995) pattern-mixture model presented in equations (2) and (3). If the data are MCAR or MAR valid estimates can be obtained with these terms excluded.

Overall et al. (1999) investigated an analysis similar to the pre-post score analysis advocated by Delucchi and Bostrom (1999), namely an endpoint analysis involving a simple change score from baseline to the last available measurement (p. 206). Their endpoint analysis is a two-stage procedure. At stage-one they obtained a simple change score from baseline to last available measurement and apply these change scores in an ANCOVA, again using pretest score on Y $\left(Y_{i1k}\right)$ and the number of available measurements for participant i $\left(J_{ik}\right)$ as covariates:

$$\left(Y_{ijk} - Y_{i1k}\right) = l_0 + l_1 J_{ik} + l_2 z + l_3 Y_{i1k} + d_{ik} .$$

Overall et al. (1999) employed pretest scores and number of available measurements as covariates because Overall et al., (1996) had shown that these covariates were necessary to control the Type I error rate in conditions where participants who drop out early show less change from the pretest than do later dropouts and completers.

Overall et al. (1999, pp. 205-209) also investigated an ANCOVA approach implemented by using PROC MIXED, though their approach differs from Wu and Bailey (1989). They included the pretest score on Y and the number of available measurements for participant i as covariates in order to have the same type of covariate control that they had in their change score analysis. Their model is

$$\begin{aligned} Y_{ijk} &= b_{0ik} + b_{1ik} t_j + e_{ijk} \\ b_{0ik} &= l_{00} + l_{01} J_{ik} + l_{02} z + l_{03} Y_{i1k} + u_{i0} \\ b_{1ik} &= l_{10} + l_{12} z + u_{i1} . \end{aligned}$$

Substituting the right hand sides of the equations for the intercept and slope into the equation for the observed data

$$\begin{aligned} Y_{ijk} = {} & l_{00} + l_{01} J_{ik} + l_{02} z + l_{03} Y_{i1k} \\ & + l_{10} t_j + l_{12} z \times t_j + u_{i0} + u_{i1} \times t_j + e_{ijk} , \end{aligned}$$

we see that pretest scores appear in the model both as dependent variable scores and as independent variable scores. As Overall et al. (1999, pp. 213-214) and Ahn, Tonidandel, and Overall (2000, pp.278-279) pointed out, use of this model has not been without controversy. A less controversial alternative is to include the pretest as a covariate, but to exclude pretest score from the dependent variable. However, simulations conducted by Overall et al. indicated that the more controversial procedure worked adequately for testing the group $\times$ time interaction.

Moreover, Ahn et al. compared the more controversial and less controversial procedure and showed that both had similar Type I error rates for testing the group $\times$ time interaction, but the procedure developed by Overall and his colleagues had better power. PROC MIXED code for the Overall et al. model is

```
proc mixed method=ml;
class id group;
model score=nrm t1 group time
time*group/solution;
random intercept time/type=un
subject=id;
```

The variable nrm is the number of measurements available for a participant. The variable t1 is the pretest score. There are three major differences between our code and theirs. First the time of last observation (nrm) is not centered. Second t1 is included in their model but not in ours. Third, the time by nrm interaction is excluded in their model.

Finally, Overall et al. (1999) investigated a two-stage ANCOVA procedure. They again used the pretest score on Y and the number of available measurements for participant i as covariates. Like the Wu and Bailey (1989) approach, Overall et al. used OLS in stage 1 to estimate the subject-specific regression coefficients. The slopes were multiplied by $t_{J_{ik}}$ and then used in a second stage ANCOVA model:

$$t_{j_{ik}} \hat{b}_{1ik} = l_{10} + l_{11} J_{ik} + l_{12} z + l_{13} Y_{i1k} + d_{ik} .$$

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 compare these methods because prior research either had not compared all the procedures just enumerated in one study under a common set of manipulated conditions, or, the comparisons were not made on all of the measures we assess. These measures are rates of Type I error and power for the test of equality of average slopes, bias in the difference in the average slopes, and the variability in estimating this difference.

Method

Nine methods of examining the group by time interaction effect in a between by within subjects repeated measures design were examined. Specifically, the methods (with their acronyms) were:

(1) the PROC MIXED analysis that presumes the data are missing at random (PMMAR),
(2) the un-weighted least squares (pattern-mixture) analysis (UWLS),
(3) Hedeker and Gibbons' (1997) approach to estimating the pattern-mixture model (HGPMM),
(4) Overall et al.'s (1999) PROC MIXED analysis that uses t1 and nrm as covariates (OPMAOC),
(5) Wu and Bailey's (1989) ANCOVA implemented in PROC MIXED (WBPMAOC),
(6) the weighted least squares ANCOVA presented by Wang-Clow et al. (1995), where the weights for the weighted least squares part of the analysis are obtained from PROC MIXED (WLSAOC),
(7) the weighted least squares ANCOVA presented by Wang-Clow et al. (1995), where the weights for the weighted least squares part of the analysis are obtained through the method of moments (See Wu & Bailey, 1998, p. 945) (WLSAOCMM),
(8) Overall et al.'s (1999) two-stage ANCOVA (OTSAOC), and
(9) Overall et al.'s (1999) two-stage endpoint ANCOVA (OEPAOC).

In the UWLS method standard errors were calculated by using the procedure presented in equation (6). However, s^2 and **D** were estimated by maximum likelihood rather than the method of moments.

We investigated two factors in our study: number of equally spaced levels of the repeated measures variable (5 and 9) and missing data mechanism (MCAR, MAR and MNAR). Overall and his colleagues (See Ahn, Tonidandel & Overall, 2000; Overall et al., 1999; Overall et al., 1996) examined the group by time interaction effect in a parallel-groups design containing a baseline score and eight additional repeated measurements; thus, for comparative purposes we had nine levels for one of our cases of number of repeated measurements. Overall and his colleagues designed their investigation to mirror design characteristics in clinical trials where a large number of repeated measurements would not be unusual. However, in behavioral science research, nine levels of the repeated measures variable may not be typical. Accordingly, we also included a smaller case, that is, five levels.

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 *i*th participant, in group *k* on the *j*th occasion:

$$Y_{ijk} = b_{0i} + b_{1i}t_j + e_{ijk}.$$

In each treatment group, data were simulated for 100 participants. The variable t_j was coded (0, 0.23077, 0.46154, 0.69231, 0.92308, 1.15385, 1.38462, 1.61538, 1.84615). To get the codes for conditions with five time points we eliminated the last four codes.

The mean for b_{0i} was 50 in both groups, implying that both treatment groups had the same population pretest mean. For Type I error data, the mean for the slope was 4.5 in treatment 1 and treatment 2 [$g_{11} = 0$, where g_{11} is defined in equation (1)], indicating identical average rates of increase over time, hence, a null condition. For our power comparisons, the slope was 9.0 in treatment 2 and 4.5 in treatment 1 ($g_{11} = 4.5$) when there were nine occasions and 12.5 in treatment 2 and 4.5 in treatment 1 ($g_{11} = 8$) when there were five occasions. The slopes for treatment 2 were selected to provide similar power for both levels of the number of occasions factor. The errors (e_{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. The variance for the residuals, conditional on time, was 240. 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 less rapidly. We also replicated the entire study changing the covariance to 12.42 from -12.42 and retaining all other features of the design. Notable differences that emerged between the two sets of conditions will be highlighted in the Results section.

Without further complications to the method, the ANCOVA methods can only be applied to participants who have at least two observations and was formulated for the situation in which the missing data occur in a monotone pattern. That is, once a participant drops out, subsequent measurements are not available. Therefore in our simulated data, every participant had an observation at the pretest and the first two follow-up occasions.

Once the data were generated, data were eliminated according to a MCAR, a MAR, or one of two MNAR missing data mechanisms. As indicated in our introduction, when the missing data mechanism is MNAR, ignoring the mechanism can result in inconsistent estimates of the unknown parameters. Accordingly, unlike Delucchi and Bostrom (1999), we compared approaches under a MCAR, a MAR, and two MNAR mechanisms. To select missing observations we used the following model

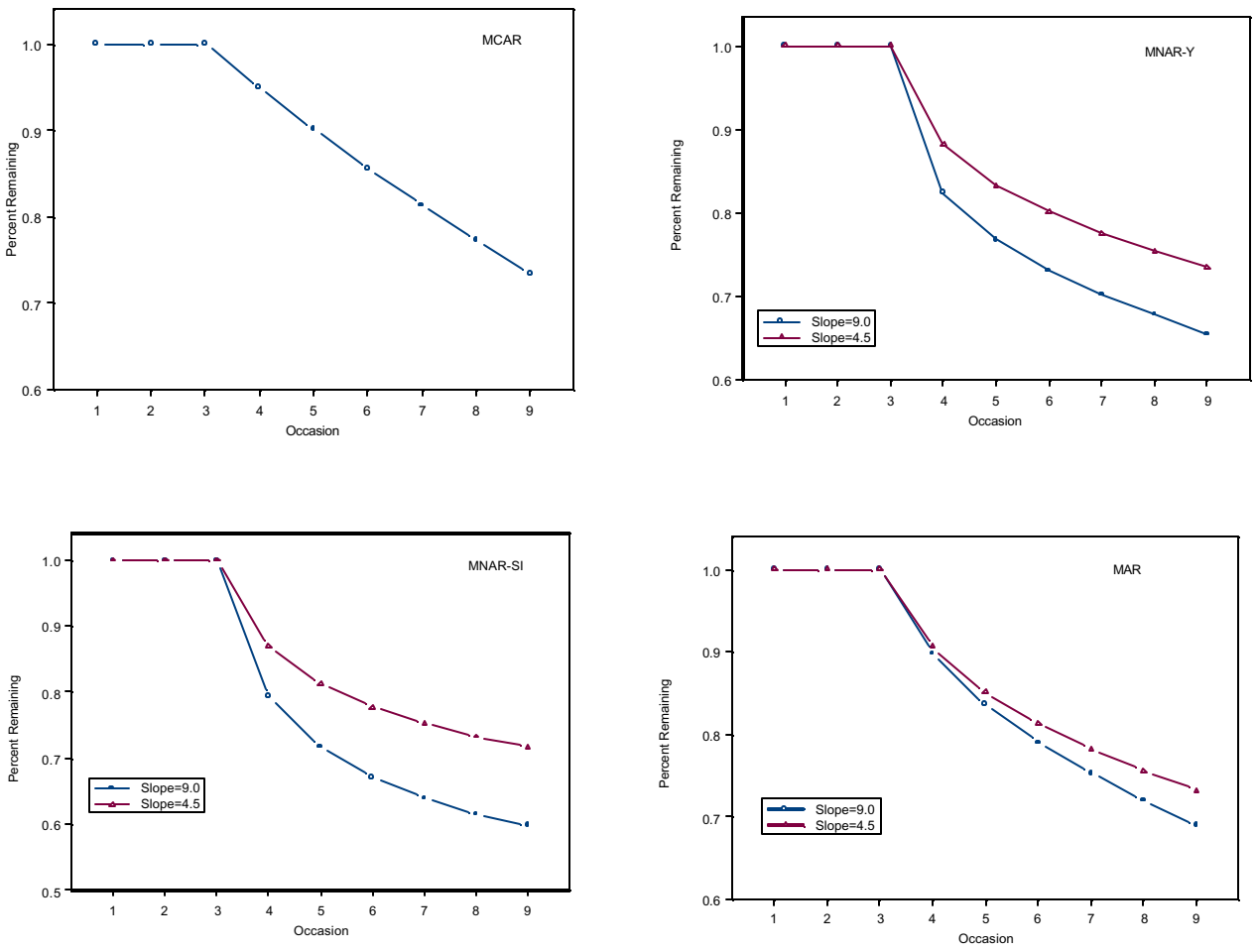
$$Z_{ijk} = q_{1j} + q_2 b_{0i} + q_3 b_{1i} + q_4 Y_{i(j-1)k} + q_5 Y_{ijk}.$$

An observation was set as missing if $U_{ijk} < f\left(Z_{ijk}\right)$ where U_{ijk} is a uniformly distributed random variable and f is the standard normal distribution. The missing data mechanism is MCAR if $q_2 = q_3 = q_4 = q_5 = 0$, MAR if $q_2 = q_3 = q_5 = 0$ and MNAR if q_2, q_3, or q_5 is not equal to zero. In one MNAR mechanism only q_2 and q_3 were not equal to zero (MNAR-SI). This mechanism meets the assumption required for the pattern-mixture model in equations (2) and (3) to yield consistent estimates. In the other MNAR mechanism, only q_5 was not equal to zero (MNAR-Y). The values of q_{1j} were selected to give cumulative missing data rates between 30% and 40% at the ninth occasion.

Figure 1 shows estimated proportions of participants remaining in the study at each occasion in the non-null condition with nine time points under the MCAR, MAR, MNAR-SI and MNAR-Y mechanisms. To obtain these estimates, 100,000 data points were generated for each treatment group. (For the MCAR mechanism, a total of 100,000 data points were generated since in our MCAR condition the dropout rate was the same in both treatments.) For our MAR condition the probability of dropping out at occasion j was positively related to the participant's score at occasion $j-1$. For our MNAR-SI condition the probability of dropping out at occasion j was positively related to the participant's intercept and slope. For our MNAR-Y condition the probability of dropping out at occasion j was positively related to the score the participant would have attained at occasion j if the participant had not dropped out. Thus in all panels of Figure 1, except the top right, drop-out rates are higher for the treatment with the average slope equal to 9 (treatment 2).

Drop out rates vary across type of missing data mechanism; however, because we will compare methods for a particular mechanism, and not the performance of a method across mechanisms, this variation in drop out rates across mechanisms is not problematic. Each condition was replicated 2,500 times. All hypotheses were conducted with a nominal alpha of .05.

Figure 1. Percent of Data that is Not Missing by Occasion and Missing Data Mechanism



Results

Tabled results are for conditions in which the correlation between the slope and intercept was negative. Important differences that emerged when the correlation between the slope and intercept was positive will be noted in the text.

Type I error rates and power are reported in Table 1 for the MCAR and MAR conditions and in Table 2 for the MNAR conditions. All procedures exhibited adequate control of the Type I error rate. However, when the missing data mechanism was MAR and the correlation between the slope and intercept was positive WLSAOCMM, WLSAOC, and WBPMAOC had higher Type I error rates than those reported in Table 1. These error rates were .067. 068, and .069, respectively, when the number of time points was five and .076, .112, and .115 for nine time points. Although in some conditions, UWLS, HGPMM, and/or OEPAOC were competitive with the other procedures in terms of power, they generally had lower power than the other procedures. Excluding HGPMM, UWLS, and OEPAOC from consideration, under the MCAR and MAR conditions, power differences were fairly small among the remaining methods. In the MCAR conditions, OTSAOC and PMMAR had the highest power estimates; in the MAR conditions WBPMAOC had the best power estimates. The slight advantage of WBPMAOC relative to PMMAR may reflect the fact that WBPMAOC resulted in treatment effect estimators with a positive bias (see Table 5) when the data were MAR, whereas, as expected

theoretically, PMMAR provided a consistent estimator of the treatment effect.

In the MNAR conditions the methods seem to separate into two groups; PMMAR, UWLS, OTSAOC, and OEPAOC tended to have lower power than the other procedures. Among OPMAOC, WBPMAOC, WLSAOC, and WLSAOCMM, WBPMAOC tended to have the highest power in MNAR-SI while WBPMAOC and OPMAOC tended to have the highest power in MNAR-Y.

The slope difference (g_{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 six methods. Table 3 contains means and standard deviations of these estimates for the MCAR and MAR conditions when $g_{11} = 0$. Table 4 contains the same information for the MNAR conditions. When $g_{11} = 0$, none of the procedures had an average estimate that was significantly different from zero. In Tables 3 and 4, UWLS and HGPMM tended to have larger standard deviations than the other procedures. The standard deviations for the remaining four procedures were similar in size.

Table 5 contains means and standard deviations of these estimates for the MCAR and MAR conditions when $g_{11} \neq 0$; Table 6 contains the same information for the MNAR conditions. Bold entries are average estimated slope differences that were significantly different from the population slope difference. The results suggest that all of the procedures are unbiased when the data were MCAR. When the data were MAR, only PMMAR did not show any significant evidence of bias. For the condition with five time points OPMAOC and HGPMM were not significantly biased. This finding probably reflects the larger standard error for the condition with five time points: For each of HGPMM and OPMAOC, the amount of estimated bias was similar when there were five and nine time points. When the covariance between the slope and intercept was positive, HGPMM exhibited more bias (average $\hat{g}_{11} = 7.680$ for five time points and $\hat{g}_{11} = 3.967$ for nine time points).

In the MNAR-SI condition, missingness depends on the subject-specific intercepts and slopes and the pattern-mixture model presented in equations (2) and (3) is expected to result in a consistent estimator of the slope difference. As expected from theory, the UWLS procedure did not result in significant evidence of bias. HGPMM, which is also intended to be unbiased under MNAR-SI, was substantially biased. In fact HGPMM exhibited the second largest amount of bias, following PMMAR. WBPMAOC, WLSAOC, WLSAOCMM were also intended to be unbiased under MNAR-SI. WLSAOCMM was unbiased and WLSAOC exhibited a small but significant bias for nine time points. WBPMAOC was biased but its bias was much smaller than that for HGPMM.

In the MNAR-Y condition missingness depends on the participant's score at occasion j; under MNAR-Y none of the procedures were expected to result in consistent estimators of the slope difference. PMMAR exhibited substantial bias for both five and nine time points. The other procedures had fairly large bias when there were five time points and less bias when there were nine time points. When the covariance between the slope and intercept was positive HGPMM was substantially biased when there were five measurement occasions; the average value of $\hat{g}_{11}$ was 7.12.

The other procedures exhibited less evidence of bias in the positive covariance case than in the negative covariance case. Although OPMAOC did not exhibit significant evidence of bias when there were nine measurement occasions and a negative covariance, OPMAOC was substantially biased when the covariance between the slope and intercept was positive with an average value for $\hat{g}_{11}$ of 4.04.

In both Tables 5 and 6 the standard deviations for UWLS and HGPMM are larger than for the other procedures which most likely accounts for their relatively poor power. The remaining procedures have similar standard deviations.

Table 1. Type I Error and Power Rates for MCAR and MAR Conditions.

Missing Data Mechanism	Method	5-levels		9-levels	
		Type I Error	Power	Type I Error	Power
MCAR	PMMAR	0.052	0.663	0.052	0.669
	UWLS	0.052	0.612	0.052	0.419
	HGPMM	0.053	0.631	0.054	0.577
	OPMAOC	0.052	0.658	0.055	0.662
	WBPMAOC	0.053	0.650	0.051	0.662
	WLSAOC	0.052	0.647	0.050	0.654
	WLSAOCMM	0.052	0.645	0.049	0.620
	OTSAOC	0.052	0.711	0.050	0.669
	OEPAOC	0.050	0.625	0.050	0.554
MAR	PMMAR	0.056	0.638	0.054	0.630
	UWLS	0.054	0.564	0.051	0.371
	HGPMM	0.047	0.555	0.048	0.473
	OPMAOC	0.055	0.645	0.053	0.645
	WBPMAOC	0.057	0.665	0.073	0.687
	WLSAOC	0.057	0.658	0.067	0.670
	WLSAOCMM	0.055	0.654	0.053	0.624
	OTSAOC	0.050	0.642	0.045	0.585
	OEPAOC	0.048	0.574	0.047	0.444

Notes: PMMAR-Proc Mixed MAR analysis; UWLS-Un-weighted least squares analysis which is ML for pattern-mixture models; HGPMM-Hedeker and Gibbons' (1997) approach to 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; OTSAOC- Overall et al.'s two-stage ANCOVA; OEPAOC- Overall et al.'s two-stage endpoint ANCOVA analysis.

Table 2. Type I Error and Power Rates for MNAR Conditions.

Missing Data Mechanism	Method	5-levels		9-levels	
		Type I Error	Power	Type I Error	Power
MNAR-SI	PMMAR	0.052	0.446	0.046	0.396
	UWLS	0.049	0.449	0.045	0.236
	HGPMM	0.053	0.364	0.044	0.273
	OPMAOC	0.056	0.531	0.048	0.505
	WBPMAOC	0.055	0.618	0.056	0.649
	WLSAOC	0.056	0.581	0.055	0.579
	WLSAOCMM	0.056	0.575	0.043	0.525
	OTSAOC	0.052	0.261	0.041	0.249
	OEPAOC	0.045	0.228	0.045	0.198
MNAR-Y	PMMAR	0.052	0.493	0.049	0.497
	UWLS	0.042	0.435	0.049	0.258
	HGPMM	0.046	0.488	0.053	0.430
	OPMAOC	0.048	0.556	0.051	0.607
	WBPMAOC	0.046	0.552	0.050	0.588
	WLSAOC	0.050	0.528	0.049	0.532
	WLSAOCMM	0.049	0.520	0.042	0.478
	OTSAOC	0.048	0.449	0.045	0.435
	OEPAOC	0.046	0.422	0.051	0.336

Notes: PMMAR-Proc Mixed MAR analysis; UWLS-Un-weighted least squares analysis which is ML for pattern-mixture models; HGPMM-Hedeker and Gibbons' (1997) approach to 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; OTSAOC- Overall et al.'s two-stage ANCOVA; OEPAOC- Overall et al.'s two-stage endpoint ANCOVA analysis.

Table 3. Mean and Standard Deviation of the Difference between the Control and Treatment Group ($g_{11} = 0$): MCAR and MAR Conditions.

Missing data		5-levels		9-levels	
Mechanism	Method	MEAN	SD	MEAN	SD
MCAR	PMMAR	0.008	3.402	-0.023	1.947
	UWLS	-0.028	3.625	-0.032	2.588
	HGPMM	-0.014	3.572	-0.029	2.150
	OPMAOC	0.005	3.408	-0.022	1.971
	WBPMAOC	0.006	3.417	-0.023	1.961
	WLSAOC	0.004	3.416	-0.021	1.967
	WLSAOCMM	0.004	3.417	-0.021	1.972
MAR	PMMAR	0.019	3.449	0.051	1.959
	UWLS	0.006	3.875	0.084	3.000
	HGPMM	0.006	3.725	0.075	2.248
	OPMAOC	0.016	3.472	0.057	1.972
	WBPMAOC	0.009	3.542	0.030	2.116
	WLSAOC	0.013	3.541	0.045	2.109
	WLSAOCMM	0.010	3.538	0.046	2.113

Notes: PMMAR-Proc Mixed MAR analysis; UWLS-Un-weighted least squares analysis which is ML for pattern-mixture models; HGPMM-Hedeker and Gibbons' (1997) approach to 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.

Table 4. Mean and Standard Deviation of the Difference between the Control and Treatment Group ($g_{11} = 0$): MNAR Conditions.

Missing Data Mechanism	Method	5-levels		9-levels	
		MEAN	SD	MEAN	SD
MNAR-SI	PMMAR	0.000	3.523	0.012	1.950
	UWLS	0.086	4.008	-0.053	3.206
	HGPMM	0.063	3.903	0.016	2.376
	OPMAOC	0.028	3.545	-0.003	1.993
	WBPMAOC	0.025	3.538	0.014	2.007
	WLSAOC	0.033	3.551	-0.013	2.037
	WLSAOCMM	0.035	3.554	-0.012	2.042
MNAR-Y	PMMAR	-0.043	3.520	-0.028	1.968
	UWLS	-0.008	3.860	-0.045	3.105
	HGPMM	-0.066	3.783	-0.024	2.351
	OPMAOC	-0.044	3.480	-0.022	1.956
	WBPMAOC	-0.046	3.482	-0.021	1.936
	WLSAOC	-0.042	3.499	-0.023	1.970
	WLSAOCMM	-0.040	3.497	-0.020	1.978

Notes: PMMAR-Proc Mixed MAR analysis; UWLS-Un-weighted least squares analysis which is ML for pattern-mixture models; HGPMM-Hedeker and Gibbons' (1997) approach to 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.

Table 5. Mean and Standard Deviation of the Difference between the Control and Treatment Group ($g_{11} \neq 0$): MCAR and MAR Conditions.

Missing Data		5-levels $g_{11} = 8.0$		9-levels $g_{11} = 4.5$	
Mechanism	Method	MEAN	SD	MEAN	SD
MCAR	PMMAR	8.036	3.357	4.501	1.895
	UWLS	8.094	3.597	4.542	2.560
	HGPMM	8.109	3.558	4.495	2.082
	OPMAOC	8.046	3.365	4.511	1.907
	WBPMAOC	8.026	3.381	4.503	1.899
	WLSAOC	8.032	3.381	4.513	1.901
	WLSAOCMM	8.033	3.382	4.514	1.902
MAR	PMMAR	8.006	3.544	4.489	1.969
	UWLS	8.253	3.993	4.805	3.031
	HGPMM	7.862	3.833	4.311	2.235
	OPMAOC	8.137	3.567	4.618	1.986
	WBPMAOC	8.374	3.645	4.888	2.124
	WLSAOC	8.338	3.644	4.865	2.113
	WLSAOCMM	8.334	3.644	4.863	2.117

Notes: PMMAR-Proc Mixed MAR analysis; UWLS-Un-weighted least squares analysis which is ML for pattern-mixture models; HGPMM-Hedeker and Gibbons' (1997) approach to 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. Bold values indicate average estimates that are significantly different than the population slope difference.

Table 6. Mean and Standard Deviation of the Difference between the Control and Treatment Group ($g_{11} \neq 0$): MNAR Conditions.

Missing Data		5-levels $g_{11} = 8.0$		9-levels $g_{11} = 4.5$	
Mechanism	Method	MEAN	SD	MEAN	SD
MNAR-SI	PMMAR	6.606	3.671	3.411	2.037
	UWLS	7.978	4.391	4.394	3.509
	HGPMM	6.992	4.344	3.660	2.541
	OPMAOC	7.489	3.676	4.057	2.052
	WBPMAOC	8.318	3.733	4.809	2.082
	WLSAOC	8.069	3.737	4.588	2.127
	WLSAOCMM	8.066	3.739	4.582	2.136
MNAR-Y	PMMAR	6.893	3.437	3.964	1.997
	UWLS	7.395	3.978	4.301	3.320
	HGPMM	7.667	3.868	4.405	2.390
	OPMAOC	7.477	3.452	4.455	1.996
	WBPMAOC	7.491	3.439	4.379	1.994
	WLSAOC	7.310	3.476	4.194	2.051
	WLSAOCMM	7.309	3.477	4.202	2.052

Notes: PMMAR-Proc Mixed MAR analysis; UWLS-Un-weighted least squares analysis which is ML for pattern-mixture models; HGPMM-Hedeker and Gibbons' (1997) approach to 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. Bold values indicate average estimates that are significantly different than the population slope difference.

Additional Conditions and Results

Our results indicate that UWLS can be inefficient and have low power. As noted earlier the sampling variance of the UWLS estimator of the slope difference is the (2,2) element of $\sum_k \sum_i \frac{\widehat{\mathbf{V}}_i}{n_k^2}$ where $\widehat{\mathbf{V}}_i = \hat{s}^2(\mathbf{X}_i'\mathbf{X}_i)^{-1} + \hat{\mathbf{D}}$ and therefore depends on the relative sizes of the contributions of $\hat{s}^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}$ and $\hat{\mathbf{D}}$. This being the case, in order to increase the generalizability of our results, we expanded our study by conducting additional simulations in which the **X** matrix used to generate the data

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}$$

rather than

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & .23 & .46 & .69 & .92 & 1.15 & 1.38 & 1.61 & 1.85 \end{bmatrix}.$$

These simulations were limited to MCAR and MNAR-SI missing data mechanisms. For the MAR and MNAR-Y missing data mechanisms in our study, it is not possible to change the initial **X** matrix without either increasing the rate of missing data or reducing the dependence of the missing data indicator on the variables in the missing data model to maintain the rates of missing data that occurred with the original **X** matrix. In either case, the change in the **X** matrix would be confounded with another feature of the data. For these simulations we used 1000 replications. All other features of the simulation were unchanged. Given that we only changed was the **X** matrix, the change simulates conducting a study over a longer time period.

In the MCAR and MNAR-SI conditions with the **X** matrix, all procedures controlled the Type I error rate well. The same result was found with the revised **X** matrix except when the covariance between the slope and intercept was positive and the data were MNAR-SI. Then WLSAOCMM, WLSAOC, and WBPMAOC had higher Type I error rates than with the original **X** matrix. The error rates were .072, .072, and .076, respectively, when the number of time points was five and .078, .083, and .084 for nine time points.

In general, with the new **X** matrix the UWLS procedure was more competitive in terms of sampling variability (see Tables 7 and 8, which contain results for the condition with a negative correlation between the slope and intercept) and thus in power. Thus, contrary to the results in Wang-Clow et al. (1995), UWLS can be reasonably efficient in some situations. Apparently, the efficiency improves as the sampling variance of the OLS estimators of the within-subjects regression model improves, as might happen when data are collected over a longer time span.

With the initial **X** matrix, UWLS was unbiased, as expected, in the MNAR-SI condition but HGPMM exhibited substantial bias when $g_{11} \neq 0$ and therefore had less power. This result also occurred with the revised **X** matrix (see Table 8).

PMMAR performed well in the MCAR condition in terms of bias and power. As expected from theory, PMMAR performed less well in the MNAR-SI condition. In particular, when $g_{11} \neq 0$, PMMAR exhibited evidence of bias and was not among the more powerful procedures. Similar results occurred with the revised **X** matrix (see Table 8).

With the initial **X** matrix, $g_{11} \neq 0$, and MNAR-SI missing data mechanisms, OPMAOC, tended to show evidence of bias, with bias ranging from 6% to 17% of the population slope difference. The bias was reduced with the revised **X** matrix, ranging from 3% to 5%. Similarly WBMAOC tended to show evidence of bias with the original **X** matrix, with bias ranging from 2% to 7%. Bias was reduced with the revised **X** matrix. In the MNAR-SI condition WLSAOC, and WLSAOCMM tended to exhibit very little bias and this was true with the revised **X** matrix also (see Table 8).

Table 7. Mean and Standard Deviation of the Difference between the Control and Treatment Group for the revised **X** matrix and $g_{11} = 0$: MCAR and MNAR-SI Conditions.

Missing Data		5-levels		9-levels	
Mechanism	Method	MEAN	SD	MEAN	SD
MCAR	PMMAR	0.017	1.486	0.075	1.386
	UWLS	0.019	1.501	0.060	1.399
	HGPMM	0.017	1.509	0.070	1.390
	OPMAOC	0.023	1.488	0.069	1.387
	WBPMAOC	0.016	1.487	0.078	1.388
	WLSAOC	0.019	1.488	0.076	1.387
	WLSAOCMM	0.019	1.488	0.076	1.387
MNAR-SI	PMMAR	0.011	1.453	-0.002	1.389
	UWLS	0.001	1.527	-0.017	1.485
	HGPMM	-0.002	1.468	0.007	1.385
	OPMAOC	0.010	1.476	-0.001	1.406
	WBPMAOC	0.008	1.494	-0.011	1.418
	WLSAOC	0.009	1.492	-0.009	1.420
	WLSAOCMM	0.009	1.492	-0.009	1.421

Notes: PMMAR-Proc Mixed MAR analysis; UWLS-Un-weighted least squares analysis which is ML for pattern-mixture models; HGPMM-Hedeker and Gibbons' (1997) approach to 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. Bold values indicate average estimates that are significantly different than the population slope difference.

Table 8. Mean and Standard Deviation of the Difference between the Control and Treatment Group for the revised **X** matrix and $g_{11} \neq 0$: MCAR and MNAR-SI Conditions.

Missing Data		5-levels $g_{11} = 8.0$		9-levels $g_{11} = 4.5$	
Mechanism	Method	MEAN	SD	MEAN	SD
MCAR	PMMAR	8.024	1.438	4.462	1.307
	UWLS	8.017	1.457	4.468	1.342
	HGPMM	8.013	1.468	4.464	1.320
	OPMAOC	8.022	1.442	4.462	1.313
	WBPMAOC	8.024	1.437	4.461	1.309
	WLSAOC	8.024	1.439	4.461	1.309
	WLSAOCMM	8.024	1.439	4.461	1.309
MNAR-SI	PMMAR	7.545	1.515	4.218	1.366
	UWLS	7.964	1.600	4.497	1.476
	HGPMM	6.999	1.621	3.867	1.413
	OPMAOC	7.751	1.533	4.304	1.378
	WBPMAOC	8.106	1.534	4.561	1.380
	WLSAOC	8.030	1.538	4.520	1.388
	WLSAOCMM	8.025	1.538	4.518	1.387

Notes: PMMAR-Proc Mixed MAR analysis; UWLS-Un-weighted least squares analysis which is ML for pattern-mixture models; HGPMM-Hedeker and Gibbons' (1997) approach to 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. Bold values indicate average estimates that are significantly different than the population slope difference.

Conclusion

The purpose of our article was to introduce and examine a number of methods of analysis for longitudinal designs in which data may be missing. Random coefficients selection models may be used to obtain estimates of parameters when data are not completely observed, that is when data are missing. As Little (1995) and others have noted, when random coefficients selection models are used, biased estimates can result if the data are MNAR and the missing data mechanism is not accounted for in the estimation procedure. An alternative method is random coefficients pattern-mixture modeling due to Little.

Little has presented a random coefficients pattern-mixture model that yields consistent estimators of the fixed effects when the missing data mechanism is MNAR-SI (i.e., the pattern of missingness is predictable from the random coefficients). Because recent evidence suggests that this pattern-mixture model can result in inefficient estimates, we presented and examined other methods of analysis that, also according to the literature, may result in better estimation of unknown parameters and which take MNAR-SI missingness into account in their analyses. In particular, we investigated methods due to Wu and Bailey (1988, 1989) and Wang-Clow et al. (1995). We also investigated several methods due to Overall et al. (1999) and we included the random coefficients selection model that ignores the missing data mechanism and an implementation of Little's pattern-mixture model that is due to Hedeker and Gibbons (1997).

All procedures except WBMAOC, WLSAOC, and WLSAOCMM controlled the Type I error rates well in all conditions. The latter three procedures had elevated Type I error rates in several conditions, although the elevation was severe only when there were nine time points. Even with nine time points, WLSAOCMM performed reasonably well, with a maximum Type I error rate of .076 for a nominal .05 test. WBMAOC and WLSAOC performed reasonably well when there were five time points with maximum estimated Type I error rates of .076 and .072 respectively.

Although no single procedure dominated the other in terms of power, WBMAOC tended to be among the more powerful procedures in all conditions. This occurred in conditions in which WBMAOC controlled the Type I error rate well in addition to the conditions in which it did not. Procedures that tended to be competitive with WBMAOC over a range of conditions were OPMAOC, WLSAOC, and WLSAOCMM.

All procedures produced estimators that were unbiased when the population treatment effect was null. Thus in the following all references to bias refer to conditions in which the treatment effect was non-null. UWLS was unbiased in MCAR and MNAR-SI conditions and had reasonably small biases in the other conditions. Consistent with evidence reported by Wu and Bailey (1989) and Wang-Clow et al. (1995), our results indicate that UWLS can be inefficient and have low power in some conditions. However, our results also indicate that UWLS can be competitive with the other procedures in terms of efficiency and power. The improved performance for UWLS occurred when the design permitted more accurate OLS estimates of the within-subject slopes. In these conditions, the standard errors produced by UWLS were fairly similar to those produced by PMMAR. Therefore a comparison of standard errors may be a useful diagnostic for determining when UWLS should be used.

HGPMM can be inefficient and have low power in some conditions though it tends to be as or more efficient that UWLS. And like UWLS, efficiency and power for HGPPM improved when the design permitted more accurate OLS estimates of the within-subject slopes. Unlike UWLS, HGPMM produced a substantially biased estimate of the treatment effect in the MNAR-SI condition.

This is a serious weakness because the pattern-mixture model is designed to be unbiased in the MNAR-SI condition. It should be noted, however, that the bias of the Hedeker and Gibbons' approach might improve if participants with different missing data patterns were combined into several missing data groups based on the similarity of the time points at which the data were missing. In addition if, within each treatment group, the expected value of the slope is the same for all participants with incomplete data, then the Hedeker and Gibbons' approach should result in an unbiased estimator of the treatment effect.

WBMAOC tended to have levels of bias similar to UWLS except with the original **X** matrix in the MNAR-SI condition. Then WBMAOC was slightly more biased. Similarly, OPMAOC also tended to have levels of bias similar to those of UWLS except in the MNAR-SI condition with the original **X** matrix. Then it tended to exhibit more bias than WBMAOC. WLSAOC and WLSAOCMM tended to have levels of bias similar to UWLS except with the original **X** matrix, nine measurement occasions, and the MNAR-Y missing data mechanism. Then WLSAOC and WLSAOCMM were more biased than UWLS, WBMAOC, and OPMAOC. PMMAR was unbiased in MCAR and MAR conditions, but exhibited fairly substantial bias in the MNAR conditions.

Our analyses of bias, sampling variability, Type I error and power indicated that no one procedure performed best for all missing data mechanisms. Clearly if one were to have valid information about the type of missing data, the information should be taken into account in selecting a procedure. Nevertheless, in our view, the Overall et al. (1999) ANCOVA (OPMAOC) performed better than the others over the range of conditions considered in the research, even though in any particular condition it may have been outperformed by one of the remaining procedures. The main drawback in OPMAOC was its negative bias in the MNAR-SI conditions; this bias made it less competitive in terms of power with other procedures, in particular with the Wu and Bailey (1989) procedure (WLSAOCMM), the Wu and Bailey procedure implemented with our PROC MIXED program (WBPMAOC), and the Wang-Clow et al. (1995) ANCOVA procedure with weights estimated using results from PROC MIXED (WLSAOC).

WLSAOCMM also tended to perform well in terms of bias, sampling variability, Type I error and power over a range of conditions. Its main weakness was a somewhat elevated Type I error rate in some conditions. However, its maximum estimated Type I error rate was .078. WBPMAOC and WLSAOC performed well when there were five time points, but showed elevated Type I error rates in some conditions with nine time points. Because these procedures tended to be among the most powerful in conditions in which they controlled the Type I error rate, they may be attractive when there are relatively few time points.

Of course, as is true of all empirical studies, the generalizability of our results is limited by the design of the study. The procedures may perform differently if different models for dropping out are adopted. Of particular interest are conditions in which the parameters for the missing data model vary across treatment groups.

References

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.

Delucchi, K., & Bostrom, A. (1999). Small sample longitudinal clinical trials with missing data: A comparison of analytic methods. *Psychological Methods*, *4*, 158-172.

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

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Inc.

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

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

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*, *1*, 377-388.

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

SAS Institute. (1995). *Introduction to the MIXED procedure: Course Notes.* Cary, NC: Author.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall/CRC,

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

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

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.